A Computational Model for the Construction of Knowledge-based Collaborative Learning Distributed Applications

PhD candidate: Santi Caballé Llobet

Thesis supervisors: Fatos Xhafa and Thanasis Daradoumis

PhD in Computer Science

Program on the Information and Knowledge Society

Open University of Catalonia (UOC)

September 17th, 2008

To those who showed me the path

Acknowlegments

I would like to thank my supervisors for their human and scientific support.

I also wish to express all my gratitude to lecturers Pere Mariné, Màrius Gómez, Santi Codolà, Jose Raya and Oriol Martí as well as 370 graduate and undergraduate students from the Open University of Catalonia who eagerly participated in the experiences reported in this thesis.

Abstract

An important research topic in Computer Supported Collaborative Learning (CSCL) is to explore the importance of efficient management of event information generated from group activity in collaborative learning practices for its further use in extracting and providing knowledge on interaction behavior. This view is especially relevant in the current shifting from a traditional educational paradigm (centered in the figure of a masterful instructor) to an emergent educational paradigm which considers students as active and central actors in their learning process. In this new scenario, students learn, with the help of instructors, technology and other students, what they will potentially need in order to develop their future academic or professional activities.

The essential issue here is first how to design a CSCL platform that can be used for real, long-term, complex collaborative problem-solving situations and which enables the instructor to both analyze group interaction effectively and provide an adequate support when needed. Secondly, how to extract relevant knowledge from collaboration in order to provide learners with efficient awareness and feedback as regards individual and group performance and assessment. The achievement of these tasks involve the design of a conceptual framework of collaborative learning interaction that structures and classifies the information generated in a collaborative application at several levels of description. Computational models are then to realize this conceptual approach for an efficient management of the knowledge produced by the individual and group activity as well as the possibility of exploiting this knowledge further as a meta-cognitive tool for real-time coaching and regulating the collaborative learning process.

In addition, CSCL needs have been evolving over the last years accordingly with more and more demanding pedagogical and technological requirements. On-line collaborative learning environments no longer depend on homogeneous groups, static content and resources, and single pedagogies, but high customization and flexibility are a must in this context. As a result, current educational organizations' needs involve extending and moving to highly customized learning and teaching forms in timely fashion, each incorporating its own pedagogical approach, each targeting a specific learning goal, and each incorporating its specific resources. Organizations' demands also include cost-effective integration of legacy and separated learning systems, from different institutions, departments and courses, which are implemented in different languages, supported by heterogeneous platforms and distributed everywhere, to name some of them.

These entire issues certainly represent a great challenge for current and future research in this field. Therefore, further efforts need to be made that help developers, technologists and pedagogists overcome the demanding requirements currently found in the CSCL domain as well as provide modern educational organizations with fast, flexible and effective solutions for the enhancement and improvement of the collaborative learning performance and outcomes. This thesis proposes a first step toward these goals.

Main thesis' contributions

The main contribution in this thesis is the exploration of the importance of an efficient management of information generated from group activity in CSCL practices for its further use in extracting and providing knowledge on interaction behavior. To this end, the first step is to investigate a conceptual model for data analysis and management so as to identify the many kinds of indicators that describe collaboration and learning and classify them into high-level potential categories of effective collaboration. Indeed, there are more evident key discourse elements and aspects than those shown by the literature, which play an important role both for promoting student participation and enhancing group and individual performance, such as, the impact and effectiveness of students' contributions, among others, that are explored in this work. By making these elements explicit, the discussion model proposed accomplishes high students' participation rates and contribution quality in a more natural and effective way. This approach goes beyond a mere interaction analysis of asynchronous discussion in the sense that it builds a multi-functional model that fosters knowledge sharing and construction, develops a strong sense of community among students, provides tutors with a powerful tool for students' monitoring, discussion regulation, while it allows for peer facilitation through self, peer and group awareness and assessment.

The results of the research described so far motivates the development of a computational system as the translation from the conceptual model into a computer system that implements the management of the information and knowledge acquired from the group activity, so as to be efficiently fed back to the collaboration. The achievement of a generic, robust, flexible, interoperable, reusable computational model that meets the fundamental functional needs shared by any collaborative learning experience is largely investigated in this thesis. The systematic reuse of this computational model permits a fast adaptation to new learning and teaching requirements, such as learning by discussion, by relying on the most advanced software engineering processes and methodologies from the field of software reuse, and thus important benefits are expected in terms of productivity, quality, and cost.

Therefore, another important contribution is to explore and extend suitable software reuse techniques, such as Generic Programming, so as to allow the computational model to be successfully particularized in as many as situations as possible without losing efficiency in the process. In particular, based on domain analysis techniques, a high-level computational description and formalization of the CSCL domain are identified and modeled. Then, different specific-platform developments that realize the conceptual description are provided. It is also explored a certain level of automation by means of advanced techniques based on Service-Oriented Architectures and Web-services while passing from the conceptual specification to the desired realization, which greatly facilitates the development of CSCL applications using this computational model.

Based on the outcomes of these investigations, this thesis contributes with computational collaborative learning systems, which are capable of managing both qualitative and quantitative information and transforming it into useful knowledge for all the implicated parties in an efficient and clear way. This is achieved by both the specific assessment of each contribution by the tutor who supervises the discussion and by rich statistical information about student's participation. This statistical data is automatically provided by the system; for instance, statistical data sheds light on the students' engagement in the discussion forum or how much interest drew the student's intervention in the form of participation impact, level of passivity, proactivity, reactivity, and so on. The aim is to provide both a deeper understanding of the actual discussion process and a more objective assessment of individual and group activity.

This information is then processed and analyzed by means of a multivariate statistical model in order to extract useful knowledge about the collaboration. The knowledge acquired is communicated back to the members of the learning group and their tutor in appropriate formats, thus providing valuable awareness and feedback of group interaction and performance as well as may help identify and assess the real skills and intentions of participants. The most important benefit expected from the conceptual model for interaction data analysis and management is a great improvement and enhancement of the learning and teaching collaborative experiences.

Finally, the possibilities of using distributed and Grid technology to support real CSCL environments are also extensively explored in this thesis. The results of this investigation lead to conclude that the features provided by these technologies form an ideal context for supporting and meeting demanding requirements of collaborative learning applications. This approach is taken one step further for enhancing the possibilities of the computational model in the CSCL domain and it is successfully adopted on an empirical and application basis. From the results achieved, it is proved the feasibility of distributed technologies to considerably enhance and improve the collaborative learning experience. In particular, the use of Grid computing is successfully applied for the specific purpose of increasing the efficiency of processing a large amount of information from group activity log files.

Contents

List of Figures

List of Tables

Introduction

Computer-Supported Collaborative Learning (CSCL) is one of the most influencing research paradigms dedicated to improve teaching and learning with the help of modern information and communication technology (Koschmann, 1996; Dillenburg, 1999a; Dimitriadis, et al., 2002; Stahl, 2002). Collaborative or group learning refers to instructional methods where students are encouraged to work together on learning tasks. As an example, project-based collaborative learning proves to be a very successful method to that end (Zumbach, et al., 2003). As such, CSCL applications aim to create virtual collaborative learning environments where students and tutors are able to cooperate with each other in order to accomplish a common learning goal.

One key issue when developing CSCL applications is interaction data analysis, a core function for the support of coaching and evaluation in CSCL environments. It relies on information captured from the actions performed by participants during the collaborative process (Dillenburg, 1999b; Martínez, et al., 2003). The efficient embedding of this information and of the extracted knowledge into CSCL applications sets the basis for enhancing monitoring (Daradoumis et al., 2004) awareness (Gutwin et al., 1995) and feedback (Zumback et al., 2005) to achieve a successful learning process in collaborative environments.

This issue is especially relevant in the context of the Bologna Process (Kulesza and Reinalda, 2006) and the current shifting from a traditional educational paradigm (centered on the figure of a masterful instructor) to an emergent educational paradigm which considers students as active and central actors in their learning process. In this new paradigm students learn, with the help of instructors, technology and other students, what they will potentially need in order to develop their future academic or professional activities. The instructor's role is, therefore, moving from one related to a knowledge transmission agent to another related to a specialist agent who designs the course, guides, assists and supervises the student's learning process (Simonson et al., 2003).

These new educational views are strongly related to well-fundamented pedagogical theories, such as constuctivism (Jonassen, 1994). Constructivism's central idea is that human learning is constructed rather than received, that learners build new knowledge upon the foundation of previous learning. This view of learning sharply contrasts with the traditional one in which learning is a passive and individual transmission of information from the instructor to learner.

In the context of these new principles and theories, the provision of relevant and selected knowledge about collaboration becomes essential in support for the instructors' monitoring tasks as well as in enhancing fundamental aspects of the learning process, such as problem-solving abilities by means of supporting peer- and self-evaluation and allowing learners to be aware of the progress of their peers and of their own.

On the other hand, over the last years, e-Learning and in particular CSCL needs have been evolving accordingly with more and more demanding pedagogical and technological requirements. Current educational organizations' needs involve extending and moving to highly customized learning and teaching forms in timely fashion, each incorporating its own pedagogical approach, each targeting a specific learning goal, and each incorporating its specific resources. Moreover, organizations' demands include a cost-effective integration of legacy and separated learning systems, from different institutions, departments and courses, which are implemented in different languages, supported by heterogeneous platforms and distributed everywhere, to name some of them (Ateyeh and Lockemann, 2006).

As a result, modern CSCL environments no longer depend on homogeneous groups, static content and resources, and single pedagogies, but high customization and flexibility are a must in this context, meaning that collaborative learning practices need to be continuously adapted, adjusted, and personalized to each specific target learning group. These very demanding needs of the CSCL domain represent a great challenge for the research community to satisfy.

This thesis proposes a innovative multi-fold approach in the form of a computational model for knowledge-centered distributed learning with the aim of meeting the current and demanding needs found in the CSCL domain. To this end, the main contributions are (i) a complete framework for an exhaustive and efficient management and analysis of the interaction data generated during CSCL practices, (ii) an advanced reuse-based software engineering methodology for developing CSCL applications in an effective and timely fashion (iii) the leveraging of distributed and Grid infrastructure to greatly enhance the effectiveness of the learning process, and (iv) an analytical multi-evaluation process of the effects of the previous approaches on the on-line group discussion dynamics and learning process in general found in the real context of the Open University of Catalonia.

The merge of these synergies represents an attractive but quite laborious challenge that will yield CSCL systems capable of providing more effective answers on how to improve and enhance the on-line collaborative learning experience as well as to achieve a more effective collaboration (McGrath, 1991; MacDonald, 2003; Sfard, 1998; Soller, 2001; Webb, 1992).

Knowledge-centered CSCL

One key issue when designing CSCL applications is the interaction data analysis, which is a core function for the support of coaching and evaluation in CSCL environments (McGrath, 1991; Webb, 1992). It relies on information captured from the actions performed by participants during the collaborative process (Dillenbourg, 1999a; Martínez, de la Fuente and Dimitriadis, 2003). Indeed, on-line collaborative learning involves a high degree of user-user and user-system interaction and hence generate a huge amount of event information. This information is an important data source for supporting group activity with relevant information as well as for understanding, explaining and predicting patterns of group behavior and thus it needs to be easily collected, represented and automatically processed by computational models in order to extract essential knowledge about the collaborative process.

The provision of effective knowledge extracted from the information collected in CSCL environments is essential for any form of cooperation, namely coordination, communication and collaboration (Ellis et al., 1991). It allows implicit coordination of collaborative learning, opportunities for informal, spontaneous communication in both synchronous and asynchronous modes. Users should be aware of the current activity in the group (the contribution of other members, their location and availability, the users working on a shared document at the same time and so on) and should know what other co-participants are doing in real time (e.g. during a multi-user editor session, who is editing and what is being shown). In an asynchronous context, users must know the activities performed by receiving deferred information of who, when, how and where others' interactions have been performed, and also why these interactions have been performed, which implies receiving complex knowledge of the interaction history.

As a result, participants in a collaborative learning experience may greatly enhance their abilities by increasing their knowledge about others in terms of cognitive processes and skills of the students and the group as a whole in solving problems, individual and group effectiveness regarding participation and interaction behavior, social support and help and so on (Dillenbourg, 1999; Daradoumis et al., 2005). From this view, the success of CSCL applications depends to a great extent on the capability of such applications to embed information and knowledge of group activity and use it to achieve a more effective group monitoring (Daradoumis et al., 2006) as well as constantly provide group members with as much awareness and feedback as possible. Awareness (Gutwin, Stark and Greenberg, 1995) refers to the knowledge provided to participants about both what other participants are doing at the same time and what they did in the past, whereas feedback (Zumbach et al., 2003) goes one step further than awareness by providing exhaustive and elaborated information and knowledge of what is going on in the group over a long period of time. Furthermore, the persistent storage of the knowledge extracted as group memory (Conklin, 1992) is essential for both students and tutors since it allows participants to access not only the latest documents and data, which are commonly stored for later retrieval, but also the context in which they were created. Group memory also allows tutors to track the collaborative learning process for several purposes such as scaffolding and assessment of the learning outcome.

Therefore, the efficient embedding of this information and of the knowledge extracted into the group activity sets the basis for enhancing monitoring, awareness and feedback about what is happening during collaboration. Indeed, it is essential for group members to be aware of others' participation in the collaborative process as this may enhance the collaboration a great deal in terms of decision-making, group organization, social engagement, support, and so on. Moreover, providing appropriate feedback about the collaborative activities may impact positively on the motivation, emotional state, and groups' well-being in on-line collaborative learning (Caballé et al., 2005b) by means of a steady tracking of parameters related to group functioning, task performance and scaffolding (Daradoumis et al., 2005) and by giving a constant feedback of these parameters to the group. Note that in this context information refers to quantitative and qualitative data generated by the learning group whereas knowledge refers to the result of the treatment of this information in terms of analysis techniques and interpretations that will be presented to the same group that generated it.

Given that the large amount of information generated during on line group activity may require much time to be processed, an effective way of collecting, analyzing and present this information is required. To this end, a successful CSCL system needs to distinguish and account for explicitly several stages of information management that overcomes important barriers (Caballé et al., 2007f), such as (i) how to process the large amount of information collected during group activity efficiently in order to facilitate its later analysis and make the extracted knowledge available to the participants even in real time, and (ii) how information should be analyzed and what kind of knowledge should be extracted to be presented to the participants in order to provide the best possible support and monitoring of their learning and instructional processes.

Achieving a clear and well-structured conceptual model can greatly facilitate the design of a computational model that implements the process of embedding information and knowledge into CSCL applications (Caballé et al., 2006). Indeed, the structuring and classification of the event information into high-level collaborative processes as well as the identification of potential mid- and low-level indicators that measure and evaluate each process, can contribute and facilitate the building of a portable, general collaborative learning ontology for the representation, learning and inference of knowledge about each collaborative process (Daradoumis et al., 2005). This allows the design of effective computational systems that reflect as accurately as possible task performance, individual and group behavior, interaction dynamics, members' relationships and group support (Caballé et al., 2007e).

Engineering CSCL systems by reusability

A generic, robust, flexible, interoperable, reusable computational model that meets the fundamental functional needs shared by any collaborative learning experience is largely expected by the research community and industry (Caballé and Xhafa, 2003). Indeed, CSCL applications are extensively used by all forms of higher education and especially in on-line distance education where open universities have a central role and use CSCL tools massively in all their formation cycles.

Due to this extensive use, CSCL becomes very attractive for domain software developers who have recently provided a number of architecture solutions (Pahl, 2007; Caballé, 2008d, Caballé, 2007e) with the aim of reusing the large number of common requirements shared by CSCL applications. Common needs in CSCL include support for three essential aspects of collaboration, namely coordination, collaboration and communication; with communication being the base for reaching coordination and collaboration (Caballé et al., 2004) in synchronous (i.e., cooperation at the same time) or asynchronous (i.e., cooperation at different times) collaboration modes. In addition, the representation and analysis of group activity interaction forms one of the paradigmatic principles of the CSCL domain (Dillenbourg, 1999a) and should form part of the very rationale of all CSCL applications (Martínez, de la Fuente and Dimitriadis, 2003). Finally, in order to improve collaboration in a group it is essential to provide measures and rules to resolve authentication and authorization issues and so protect the system from intentional or accidental ill use as well as to perform all the system control and maintenance for the correct administration of the system.

Generic platforms, frameworks and components are normally developed for the construction of complex software systems through software reuse techniques, such as Generic Programming, Domain-based Analysis, Feature Modeling, Service-Oriented Architecture, and so on (Czarnecki and Eisenecker, 2000; Bacelo, 2002; Caballé and Xhafa, 2003; Gomaa, 2005; Ateyeh and Lockemann, 2006). Indeed, in the context of generic architectures and platforms, software reuse is by far one of the main concerns in the software industry and it is increasingly recognized its strategic importance in terms of productivity, quality and cost (Czarneki, 2005). However, despite the advance in software reuse, reuse capacity is still in an incipient status, mainly due to the short in scope of the reuse techniques such as classes, components, and frameworks, also so-called ”reuse in the small”. There is, therefore, a need for increasing the level of reuse by extending the scope and, as a consequence, the impact on the software development, also so called ”reuse in the large” (Ateyeh and Lockemann, 2006). This is chiefly fulfilled by extracting the commonality and variability features of systems given a specific, wide domain and then reusing them for the construction of single systems in the same domain (Gomaa, 2005). Thus, neither longer is necessary to ”reinvent the wheel” nor to develop a new system from scratch. This way, organizations can consolidate and adapt their existing key software assets to meet the ever changing requirements and needs. These approaches have been successfully applied to different domains thus providing cost-effective interoperable applications of increased quality in timely fashion. The rapid change and evolution of requirements in the CSCL domain raises new challenges to software developers, who in turn demands more powerful reuse-based software techniques that provide more flexible, adaptable, modular, and maintainable software.

Therefore, leveraging the latest software reuse principles, a generic service-oriented component-based computational model in the collaborative learning context is intended to form the very rationale of complex CSCL environments in a wide range of learning situations and pedagogical goals (Caballé et al., 2007e). As a result, domain developers can derive specific CSCL applications by systematically adapting and tailoring this reusable computational model for the construction of effective, affordable and timely newly CSCL tools, which are modular, flexible, interoperable and maintainable, and a fast adaptation of existing applications to newly learning and teaching requirements (Caballé et al., 2004; Ateyeh and Lockemann, 2006).

Distributed and Grid infrastructure for CSCL

In addition to the fundamental needs found in the CSCL domain, modern, complex on-line collaborative learning environments must provide both advanced enablement for distributed collaborative activities and the necessary functionalities and resources to all participants regardless the location of both participants and resources. The aim is to enable the collaborative learning experience in open, dynamic, large-scale and heterogeneous environments (Pankatrius and Vossen, 2003). To help achieve this aim, Grid computing (Foster and Kesselman, 1998) has emerged as a way of capturing the vision of a networked computing system that provides broad access not only to massive information resources, but to massive computational resources as well (Xhafa et al., 2004). The concept of computational Grid has its origins in wide-area distributed computing, and extends to a large-scale, flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources.

Therefore, a computational model for CSCL may greatly enhance its possibilities by taking advantage of the inherent performance potential of distributed technologies, such as Grid and peer-to-peer (P2P) (Amoretti et al., 2005). Furthermore, a set of interesting non-functional features may be provided by these distributed technologies, which represent an ideal context for supporting and producing major benefits for CSCL applications. Such important features include (Pankatrius and Vossen, 2003, Caballé et al., 2004): enable and scale the involvement of an increasing large number of single/group participants (teachers, students, tutors among others) who can be widely separated by geography and/or time, possibly situated at very different locations, and transparently share a huge variety of both software and hardware resources, even in real time. The combination of distributed technologies with service-oriented architectures (SOA) allows developers to cope with essential issues in CSCL, such as integration, interoperability reliability, and flexibility so as to meet the needs of different, heterogeneous and legacy environments (GuiLing et al., 2005). To this end, a computational platform for developing CSCL applications is to consider the non-functional requirements found in this domain in a transparent manner by leveraging the latest software development methodologies, and especially the service-oriented approach. The ultimate goal of CSCL computational platforms is to serve as mediators between the application layer and the infrastructure layer making CSCL tools independent from the type and evolution of the underlying infrastructure.

Further demanding non-functional requirements found in CSCL include the provision of effective information and knowledge about what is occurring in the group activity. This implies receiving knowledge simultaneously both synchronously and asynchronously since the current and history interaction data shown are continuously updated. However, CSCL applications are characterized by a high degree of user-user interaction and hence generate a huge amount of information usually maintained in the form of event type information. In addition, users are continuously interacting with the system (creating documents, reading others' contributions, etc.) thus generating a lot of events, which, once collected, they must be classified, processed, structured and analyzed (Caballé et al., 2005). In order to capture the interaction correctly, this event information should be classified into different categories such as work sessions, messages, workspaces, documents and many other objects and thus may generate large size of information, especially in real online collaborative learning that comprise complex learning activities to be carried out during a rather long period of time and involve a considerable number of participants. This information may also include a great variety of types and formats and hence tends to be large in size (Caballé et al., 2005).

Therefore, the supply of efficient and transparent knowledge to users in this context is a significant challenge. Moreover, the need to make the analyzed information available in real time entails that processing requirements may not be met with a single computer. Indeed, the lack of sufficient computational resources is the main obstacle for processing large amounts of data log files in real time. In real situations this processing tends to be done later, after the completion of the learning activity, thus having less impact on it (Xhafa et al., 2004). Recently, Grid technology is also increasingly being used in this context to reduce the overall, censored time in processing data by taking advantage of its large computing support. Thus, in this thesis, a Grid approach is used to match the time processing requirements in order to make the processed information available to the group members in an efficient manner, to compute statistical results and to present the results to the group members and tutors, who are in different locations, as a means of facilitating the group activity, decision making, task accomplishment, and assessment of the progress of the group etc.

Furthermore, Grid technology provides a set of interesting features which represent an ideal context for supporting and producing major benefits for CSCL applications (Foster and Kesselman, 1998; Amin et al., 2002). Such important features include: large scale of Grid infrastructures, wide geographical distribution of resources, multiple administrations from different organizations, transparent and dependable access as well as the capability of granting access to shared, heterogeneous resources in very dynamic environments (Bote-Lorenzo et al., 2007). Considering the benefits provided by Grid it is possible for educational organizations to make use of true collaborative learning environments that enable the involvement of large number of single/group participants (teachers, students, tutors among others) who can potentially belong to many different organizations, possibly situated at very different locations, and transparently share a huge variety of both software and hardware resources while enhancing human-to-human interaction (synchronously or asynchronously) through a friendly complex user interface.

To sum up, leveraging the inherent performance potential of distributed and Grid infrastructure for CSCL applications makes it possible to greatly enhance the collaboration between users in terms of both participant scalability (adding as many participants/groups as necessary) and resource availability (replicating and executing them in multiple Grid nodes) enabling collaboration as the most important learning method. In addition, distributed infrastructure may help overcome important barriers in the form of certain non-functional requirements, such as fault-tolerance, performance, and interoperability, which are especially frustrating when they are not fulfilled appropriately during the collaborative learning activity. Indeed, these requirements may have considerable repercussions on the learning performance and outcomes as their lack impedes the normal learning flow as well as discriminates learners in terms of technology skills and technical equipment (Caballé et al., 2007g).

Evaluation on real CSCL practices

In order to correctly evaluate the computational models for the development of effective CSCL applications, evaluation should be considered as a process or a collection of processes to obtain and analyze significant information that support judgments of character on an object, a phenomenon, process or an event as an aid for an eventual decision on itself (Latorre et al. 1996). Evaluation in CSCL means to deal with many educational theories, each considering their own means and questions. This means that there is no a unique vision on how to deal with evaluation in CSCL. Literature shows different ways to successfully cope with evaluation, such as surveys, interviews, empirical experimentation, case studies and computer generated statistical measurements (Manson, 1992).

The Open University of Catalonia(UOC) (http://www.uoc.edu) offers distance education through the Internet since 1994. About 45,000 students, lecturers and tutors are involved in 600 on-line official courses from 23 official degrees and other PhD and post-graduate programs. The UOC has shown a strong need to monitor and evaluate real, long-term, complex, collaborative problem-solving situations and thus it represents and ideal context for experimenting and evaluating the different approaches addressed in this thesis. Thus, the needs for exhaustively evaluating the possibilities offered by a computational model for enhancing and improving the learning experience are to be met by leveraging the real collaborative learning processes and dynamics found in the virtual classrooms of the UOC. This way, the impact and repercussions of using a CSCL computational model on the collaborative learning process can be empirically reported by means of a set of surveys, user interviews, statistical measurements, and so on, involving in the evaluation process all the actors participating in the real learning experience at the UOC.

Thesis objectives

The main contribution in this thesis is four-fold. First, a conceptual model will be proposed for collaborative learning that facilitates a continuous monitoring of the learning activity, providing group members with appropriate support, as well as awareness and feedback about what is happening during collaboration. To that end, this conceptual model will include both a set of indicators that will structure and classify the information generated in the group activity at several levels of description and a process to efficiently embed this information and the knowledge acquired into CSCL applications.

Second, the indicators and the process of interaction data analysis and management defined in the conceptual model will be translated into a computational model of collaborative learning interaction. The development of this computational model, and its later enhancement by means of leveraging the inherent powerful features of distributed technologies, will be used as both a skeleton for providing the essential needs to any CSCL environment and for the achievement of an effective collaborative learning experience.

Third, for the purpose of achieving a computational platform as generic, reusable, flexible, and interoperable as possible, innovative software reuse engineering processes will be proposed to enable an effective and systematic reutilization of its generic features to facilitate the whole life cycle development of the CSCL applications using this platform.

Finally, a full-featured structured discussion forum as well as other minor CSCL applications will be developed using the computational model for validation purposes. Evaluation will come from the analysis of the repercussions in the real collaborative learning experience of these applications by involving tutors and learners that participate in the collaborative learning experience. Lessons and experience gained will be reported in using this application at the UOC for the support of real group activity.

These contributions are concreted in the following objectives:

- Identify a set of indicators that permits the collection and classification of the different types of information generated in the group activity such as task performance, group functioning and scaffolding, as well as to identify the real skills and intentions of participants

- Design a process for the efficient collection, processing, analysis, and presentation of the information and knowledge acquired from the group activity into CSCL applications.
- Define a conceptual model that formally describes and integrates both the potential indicators of effective collaboration and the process of information and knowledge management in order to constantly support group participants by providing them with as much awareness and feedback as possible as well as for assessment and monitoring purposes.
- Develop a computational system in the form of a generic specification as a translation from the conceptual model specially used to embed information and knowledge in an efficient manner.
- Elaborate a software engineering process to semi-automatically generate the realization of specific-platform implementations from independent specifications so as to greatly facilitate the reuse of the computational model.
- Enable the computational model as a generic, reusable, service-oriented platform for the systematic and flexible construction of any application in the CSCL domain.
- Explore the possibilities of distributed computing as an effective infrastructure to greatly enhance the performance of the computational model.
- Develop a complete and complex structured discussion forum that may provide new opportunities to learning methodologies so as to validate the capabilities offered by the computational model.
- Systematic construction of several simple CSCL applications and tools using the computational model with the aim at validating its reuse capacity.
- Report the experience of deploying, installing and running the discussion forum application in the real learning context of the UOC so as to support the group activity in several distance courses.
- Elaborate field tests, surveys, and user interviews involving the actors of the group activity in order to evaluate and compare the impact on the real learning experience supported by the computational model.

Thesis hypothesis

The mentioned objectives lead to formulate five sequential hypotheses in this thesis.

1. The achievement of a conceptual model for analyzing and managing data from CSCL applications should provide the means of correctly classify the many different variables that characterize collaborative interactions as well as the identification and measurement of these variables in terms of the user and system specific actions.

2. The translation from this conceptualization into a computational model should efficiently embed the treatment of the information collected from the group activity, and the later analysis and presentation of essential knowledge acquired into the collaborative process. This should result in both a positive impact on the group performance and outcomes by increasing learners' knowledge of each other and of the group activity, and the improvement of the teaching experience by facilitating the monitoring and assessment processes.

3. The innovation in software reuse in the construction of a CSCL computational platform should meet the fundamental needs of CSCL environments in terms of both functional and no functional requirements as well as the provision of extensive support and semi-automation to the development life cycle of the specific CSCL applications using this platform.

4. The use of distributed technologies to support CSCL applications in real, collaborative learning environments developed by this platform should greatly enhance and improve the collaborative learning experience, in terms of the non-functional features of the collaboration.

5. The evaluation processes designed to analyze the impact of the computational model on the real collaborative learning and teaching activity should provide appropriate quantitative and qualitative data to report and verify the feasibility and possibilities offered by this platform.

Chapter 1

Related Work

In this chapter, an overview of the literature involved in this thesis' research lines is presented. This overview will serve as background and further reference for the rest of chapters.

1.1 Knowledge-centered CSCL

When developing Computer-Supported Collaborative Learning (CSCL) environments that support online collaborative learning, several issues must be taken into account in order to ensure full support to the online learning activity. One such key issue is interaction management and analysis to support awareness, coaching and evaluation, based on information captured from the actions performed by participants during the collaborative process (Dillenbourg, 1999a).

The success of CSCL applications depends to a great extent on the capability of such applications to embed information and knowledge from the group activity's interaction and use it to achieve a more effective evaluation of collaborative learning (Dillenburg, 1999b; Avouris et al, 2003; Martínez et al., 2003; Daradoumis et al., 2005, Caballé et al., 2006). This strategic line of research is still in an incipient status (Martínez et al., 2003) as questions related to the information and knowledge embedding have not been sufficiently investigated.

An initial approach (Zumbach et al., 2003; Zumback et al., 2005) considers the use of feedback in on-line learning and its effects on group activity in general. To this end, some types of information generated by the group activity are considered as relevant knowledge to be communicated to the group members for feedback purposes but the process of how to collect the information, analyze it and extract the desired knowledge is not provided.

An automatic collection of information about interaction is presented in (Martínez et al., 2003), for the support of the formative evaluation processes and scaffolding. To this end, they present a flexible, efficient, and generic model of collaboration to be applied in different CSCL situations. Then, they represent this model by a XML-based computational model of collaborative action. Despite the fact the model distinguishes different types of interactions coming from distinct data sources, no further research is provided of how to classify, process, and present this information and the knowledge acquired in an efficient manner so as to be effectively fed back to the collaboration at the same time it is generated.

In Daradoumis (2004) and Daradoumis (2005), it is shown an integrated approach involving qualitative, quantitative and social network analysis for analyzing and assessing the performance of on-line learning groups from the information gathered during the long-term, complex, collaborative learning experience. This approach identifies several important high-level processes of collaborative learning that allows an initial classification and evaluation of the group activity interaction in terms of the learning product, group functioning, and social and help support (i.e. scaffolding):

- Learning product refers to task skills and knowledge acquired by each member as well as the quality of group work.
- Group functioning considers the analysis and assessment both of the interaction behavior of group members and the social aspects of group work.
- Scaffolding refers to social support among members as well as to task- or group functioning-oriented help provided to a participant who is not quite able or ready to achieve a task on his or her own.

However, the high-level view presented in this approach does not provide much help identify and precise the exact meaning of all learning processes and the low-level interactions occurring in the day-to-day virtual classrooms. Furthermore, the whole study was based on the BSCW groupware environment (Bentley, 1997), which shows a serious lack of tools for collecting and analyzing the appropriate information, especially in real time, and thus it was done after completion of the task, which impacts much less on the group activity (Xhafa et al., 2004).

A great deal of recent research has been done (Schellens and Valcke, 2006; De Wever, et. al., 2006; Strijbos et al., 2006; Pena-Shaff and Nicholls, 2004; Schrire, 2006; Bratitsis and Dimitracopoulou, 2006; Hew and Cheung, 2008; Puntambekar, 2006) to identify and model precise processes and indicators according to real skills and intentions shown by the participants, especially in the context of collaborative activities, such as group discussions, which are the exemplification of collaborative learning activities this thesis uses and is interested in for empirical purposes. Next, an overview of current research on how to model discussion interaction and its later processing and analysis is presented.

1.1.1 Modeling user interaction

Schellens and Valcke (2006) investigated whether collaborative learning in asynchronous discussion groups results in enhancing academic discourse and knowledge construction. Their research work showed that students in the discussion groups were very task-oriented and that higher proportions of high phases of knowledge construction were observed. Moreover, significant increases in the cognitive interaction, task-orientation and higher phases of knowledge construction were detected.

Furthermore, an important issue raised in collaborative learning interactions is the change from divergence to shared understanding and to possible construction of knowledge. The point is to understand how collaborative interactions develop over time: whether students raise new issues (ideas) more frequently as they become more familiar with the discussion and discussants, and whether shared knowledge building becomes richer over time, and subsequent evidence that students were able to construct their own understanding based on their interactions with others (Puntambekar, 2006). To this end, the model annotates and examines a variety of elements that contribute to the understanding of the nature of the collaborative interactions, such as the students' passivity, proactivity, reactivity as well as the effectiveness and impact of their contributions to the overall goal of the discussion.

Large amounts of information data are generated from asynchronous discussion which includes complex issues of the collaborative work and learning process (e.g., group well-being (McGrath, 1991) as well as self, peer and group activity evaluation (Daradoumis et al., 2006)). On the one hand, quantitative information can be managed by applying a structured process where the users' interactions are tagged with certain indicators according to a collaborative learning conversation skill taxonomy (Soller, 2001) that models the various types of interactions at different levels. Moreover, typical quantitative indicators about the participants' performance and dynamics (e.g., number of contributions written and read by each participant) are also considered as relevant to model the group functioning and task performance (Daradoumis et al., 2006).

Indeed, quantitative content analysis has been increasingly used to surpass surface level analyses in CSCL (e.g., counting messages) and several content analysis schemes have been employed to analyze transcripts of online asynchronous discussion groups in formal educational settings (De Wever, et. al., 2006). Although this research technique has been often used, standards are not yet established. As a consequence, the empirical base of the validity of the instruments is limited. Several open questions still exist, especially as concerns the unit of analysis and segmentation procedure to be followed (Strijbos et al., 2006). In a different study, a content analysis scheme has been applied to analyze the way online peer tutoring (conducted by fourth-year students) supports asynchronous discussion groups of first-year students (De Smet, et. al. 2008). This study demonstrates the important role that tutoring plays in online asynchronous discussions, which is taken into account and constitutes a contributing element of the model.

On the other hand, qualitative information is also valuable to complete the labored task of interaction analysis and evaluation of contributions. Pena-Shaff and Nicholls (2004) used a mixed approach to analyze student interactions and meaning construction in computer bulletin board discussions. Quantitative analysis was used to examine participation and interaction rates, and qualitative procedures were used to analyze knowledge construction processes and to refine a category system of indicators and descriptors. Results showed that students engaged in a knowledge construction process that was characterized chiefly by clarification, elaboration, and interpretation.

Moreover, a study by Schrire (2006) applies a merging of quantitative analysis within a qualitative methodology to build a model for the analysis of collaborative knowledge building in asynchronous discussion. The model allows examination of the communication from the multiple perspectives of interaction, cognition and discourse analysis. Content analysis of the discourse was done at a number of levels, focusing on the discussion forum itself, the discussion threads, the messages, and the exchanges and moves among the messages. As a result, it was possible to build a scheme for assessing knowledge building in asynchronous discussion groups. The scheme integrates the interactive, cognitive and discourse dimensions in CSCL. Similarly, Bratitsis and Dimitracopoulou (2006) analyze the quality of group interactions in asynchronous discussion by means of a multi-indicator model based on quantitative aspects of the active and passive behavior of participants (i.e., number of messages written/replied and read). This may help tutor infer problematic situations occurring during the learning process as well as identify individual behavior that may influence collaboration, such as passivity and arrogance.

In a more recent work, Hew and Cheung (2008) report a qualitative study examining the facilitation techniques used by student facilitators to attract their course mates to participate in asynchronous online discussions. This study differs from previous ones in the sense that it does not focus on the role of the tutor as facilitator and promoter of student participation, but it explores peer facilitation. To explore the extent to which student participation in an online discussion forum is successful, the study looks at the depth of discussion threads. Finally, it reports the facilitation techniques that were exhibited by the student facilitators. However, the mere consideration of the depth of discussion threads does not guarantee by itself the quality of the discussion; students' postings can be simply driven by socialization reasons and not directly linked to the development of the learning tasks.

1.1.2 Managing interaction data

All this information can be easily collected and automatically processed and analyzed by computers as a quantitative and qualitative data source and presented to the participants in order to provide effective information, such as how all participants are actually performing during the discussion and the dynamics of each participant with respect to the group. Consequently, the efficient embedding of all this information and of the extracted knowledge into CSCL applications sets the basis for enhancing support (Puntambekar, 2006), awareness (Gutwin, Stark, and Greenberg, 1995) and feedback (Zumbach et al., 2003) to achieve a successful discussion process in collaborative environments. Indeed, the constant and fast processing and presentation of this quantitative and qualitative data as well as their systematic analysis based on principled indicators that measure the type and the degree of group members' participation, may positively impact on participant's motivation, emotional state and problem-solving abilities and as a result enhance the acquisition of knowledge performance (Daradoumis et al., 2006).

The ultimate aim is to extract relevant knowledge of the discussion process from all possible sources (e.g., users' activity, passivity, and effectiveness; participation impact; qualitative assessment, etc.). Note that in this context information refers to quantitative and qualitative data generated by the learning group whereas knowledge refers to the result of the treatment of this information through analysis techniques and interpretation. This knowledge will be fed back and presented to the learning group members and its tutor for awareness, feedback, and scaffolding purposes.

1.2 Engineering software CSCL applications

Over the last years, a great amount of full-featured e-learning systems have appeared in the marketplace offering designers and instructors adaptative, powerful user-friendly layouts for the easy and rapid creation and organization of courses and activities, which can then be customized to the tutor's needs, learners' profile and specific pedagogical goals. In this section, an overview of existing e-learning systems for group activity purposes is first presented. Then, an overview of the latest reused-based software engineering techniques are provided provided both to meet the common needs found in the e-learning domain for reuse purposes and to systematically address the current demanding and changing requirements in this domain.

1.2.1 Groupware learning systems

Groupware is a software paradigm designed to facilitate collaboration among groups of people (Ellis et al., 1998). It has been successfully applied to CSCL domain due to the strong need for providing support to facilitate knowledge building in learning environments. In order to effectively support collaborative learning, groupware systems include a variety of collaboration tools such as document repository, virtual classrooms, whiteboards, chats, threaded discussion forums, etc. Furthermore, the World Wide Web, with a lightweight and extensible client-server architecture, allows client implementations for all popular computing platforms and an existing user base numbered in millions (Bentley, 1997). In particular, the fact that web pages present a cross-platform, uniform, simple, and intuitive interface defined by the browser results in web protocols and applications being increasingly used by organizations to manage the internal distribution of information, which encourages their extension to groupware functionality. As such, web technology forms an excellent basis for the realization of the most well-known learning groupware systems over the Internet, such as BSCL[1], and Synergeia[2].

Although most of these groupware systems support many aspects of the CSCL domain, they do not entirely contemplate the users' fundamental needs for collaborative learning environments, such as dynamic support to group awareness, specific components for awareness management, and interoperability between different applications to support collaborative work. Moreover, most of them provide neither log file processing nor tools for analyzing the processed information, and record the interaction as an ad-hoc huge amount of ill-structured information with a high degree of redundancy that requires an efficient data processing system in order to analyze this complex information. Finally, no analysis of the real students' skills and intentions are provided from the rationale of a specific conceptual model for data analysis and management and as result the awareness and feedback capabilities are strongly restricted.

Learning Management Systems (LMS) are software packages to enable the management of educational content and also integrate tools that support most of groupware needs, such as e-mail, discussion forums, chat, virtual classrooms, and so on (Baloian, 2004). Over the last years, a great amount of full-featured web-based LMS systems have appeared in the marketplace offering designers and instructors generic, powerful user-friendly layouts for the easy and rapid creation and organization of courses and activities, which can then be fleshed out with specific course materials and customized to the instructor's needs and pedagogical goals. Most of these systems take advantage of important standards such as Open

[1]BSCL is found at http://bscl.gmd.de (web page as of April 2008)
[2]Synergeia is found at http://www.synergeia.org (web page as of April 2008)

Knowledge Initiative[3] (OKI), and the IMS Global Learning Consortium[4] so as to enable sustainable interoperability with heterogeneous systems in the e-Learning domain and also in different domains. Representative LMS systems are WebCT and Blackboard[5], which are proprietary, and open-source such as Moodle[6], Sakai[7], and Interact[8], among many others.

Finally, open-source Content Management Systems (CMS) are seen as an abstraction or superset of LMS and are becoming very popular in building Web-based learning systems. Indeed, the flexibility of these systems and the powerful support from the open-source community may make CMS a better alternative to LMS. In the context of this thesis, LMS systems are certainly too restricted and impose many barriers when it comes to incorporate an unforeseen aspect in their internal design, such as the proposed conceptual model of information and knowledge. On the contrary, CMS systems have not conceived for learning purposes only and so they are open to build very flexible learning systems (Aspeli, 2007). Representative CMS are Plone [9] and its underlying application server Zope [10].

Despite the great support of LMS systems to important areas such as communication, collaboration and assessment, little support is provided in general to awareness and feedback, which is fundamental in this context. In spite of the source code of some of the LMS systems may be available, neither the adaptation of their internals (i.e. Application Programming Interface) to include news indicators to be collected from the participants' interaction is supported nor the presentation to participants in suitable formats of the knowledge acquired in a personalized, dynamic form switching to awareness, feedback, monitoring, and so on. Another common drawback is the lack of interoperability thus making the applications dependent from the programming language, underlying infrastructure, and so on.

Two LMS platforms especially have showed up in the marketplace and are being extensively adopted by educational organizations. Moodle together with the Sakai Project are the major open source movements increasing their share in the educational space. Moodle is designed using sound pedagogical principles such as constructivism, to help educators create effective online learning communities, while Sakai is a huge community source software development effort to design, build and deploy new collaborative learning environments for higher education. This allows educational institutions to highly customize Sakai to suit their pedagogical needs, and technological requirements.

In spite of providing an extensive API, at the present time there is no proper documentation in both Moodle and Sakai systems and the adaptation to different needs of important modules, such as the logging component, is rather chaotic. Furthermore, as Moodle is entirely written in PHP and Sakai in Java, this brings interoperability problems, which may influence the decision of choosing the technological platform under which it runs. In order to overcome these problems, both projects are developing a web-service API, but just a set of user and course management and administration methods have been created so far for testing purposes only.

[3]OKI is found at http://www.okiproject.org (web page as of April 2008)
[4]IMS is found at http://www.imsglobal.org (web page as of April 2008)
[5]WebCT and Blackboard have recently merged. More information at http://www.webct.com (Web page as of April 2008)
[6]Moodle is found at http://moodle.org (Web page as of April 2008)
[7]Sakai Project is found at http://www.sakaiproject.org (Web page as of April 2008)
[8]Interact is found at http://www.interactlms.org (Web page as of April 2008)
[9]Plone is found at http://www.plone.org (Web page as of April 2008)
[10]Zope is an open source application server for building content management systems, intranets, portals, and custom applications. Zope is found at http://www.zope.org (Web page as of April 2008)

Special case is also the .LRN system[11], which provides a specific user tracking module for a fully support monitoring and awareness during the collaboration, so that it is possible to present many views in the form of statistics of what is happening in each workspace, site and community. Despite this wide support and the great amount of information collected, there are neither specific indicators to classify and measure the real skills and intentions nor a process to continuously process and analyze this information generated in an efficient manner for awareness and feedback purposes. In order to fill this gap, this thesis contributes by means of the computational model in the context of the MOSAICLearning Project[12].

In overall, if we ask ourselves how users and resources from different LMS and CSCL systems are managed and what individual features make the difference when moving from one system to another, we can see that their realization and, most importantly, their implementation using specific programming languages and platforms are very different. Even so, all these systems' requirements and design are not so different from each other meaning that there exists a potentially important common basis among all of them, which avoids new systems being developed from scratch (Caballé et al., 2004). Next, a generic reused-based approach that follows these principles is provided.

1.2.2 Software engineering techniques

In this subsection, several software engineering techniques are analyzed from the view of generic architecture solutions for e-learning systems. These techniques are described within a historical perspective, from the most traditional to the latest techniques. Also, certain important contributions to the e-learning domain using these techniques are discussed.

Components have been one of the most successful resources in the software reuse field that has reached most of domains. It describes all activities in the context of a complete software life cycle on the basis of components (Ateyeh and Lockemann, 2006; Bacelo, 2002). A software component has contractually specified interfaces and explicit context dependencies can be deployed independently, and is subject to composition by third parties (Szyperski, 2003). In the literature (Dimitriadis et al., 2002), however, the application of components in the development of CSCL applications shows the reuse capacity achieved by components is clearly low due to its short of scope (Czarnecki and Eisenecker, 2000; Gomaa, 2005). These approaches also lack of crucial aspects the principle the CSCL paradigm, such as the analysis of interaction data and the presentation of the knowledge extracted.

(Anido-Rifón et al., 2001) propose a three-layered component-based framework focused on interactive and collaborative educational applications. In this approach, events produced by a given component are handled and delivered to remote components in the same application. Another component allows the sharing and distribution of events performed on the shared application user interface where user actions will be forwarded to every other user in the group. Although the proposed framework supports most scenarios taking into account their particularities, it fails to focus neither on processing and analyzing the event information from the user actions nor on how to present this information to users in terms of dynamic support to group awareness, specific components for awareness and feedback management.

[11].LRN is found at http://www.dotlrn.org (Web page as of April 2008)
[12]MOSAICLearning Project is found at http://www.mosaic.gast.it.uc3m.es (Web page as of April 2008)

(Bacelo et al., 2002), introduces an interesting approach to component-based architecture to support collaborative application designs, especially communication, cooperation, and information sharing aspects are considered. However, aspects such as awareness and knowledge management are omitted and need to be taken into account.

For the purpose of this thesis, it is worth mentioning here some approaches such as Bote et al. (2007) and Pankatrius and Vossen (2003), which point to the use of distributed computing environments in the development of components for collaborative learning domains. In the context of this research, distributed computing is also used for specific collaborative learning applications to meet many important needs of these applications, such as group activity data analysis and management, in a highly effective manner. For instance, Grid computing (Foster and Kesselman, 1998) offers high-throughput and data-intensive computing, which greatly facilitate the process of embedding information and knowledge into these applications making it possible to provide users with constant real-time awareness and feedback.

Marquès et al. (2006) also propose a distributed and decentralized infrastructure which has the aim of supporting distributed group learning and team work activities. This infrastructure is based on event distribution mechanisms providing awareness so that participants can be notified and thus be made aware of the progress of the groups they belong to. It describes how to collect and propagate events so as to notify group participants about the activities of others. However, it does not describe how this information should be presented to participants and thus some of the effectiveness in providing awareness is lost.

Although all these approaches have been successful for the construction of effective groupware and specifically CSCL applications, new engineering software paradigms and techniques are needed to address the new demanding pedagogical and technological requirements appearing in CSCL domain while reducing the overall engineering effort required to meet them. To this end, generic platforms and frameworks are normally developed for the production of complex software systems through the reuse technique by focusing on family systems rather than single systems (Czarnecki and Eisenecker, 2000).(Czarnecki, 2005). This approach has been successfully applied to different domains thus providing applications of increased quality reducing both cost and development time. To this end, Generic Programming (GP) paradigm (Czarnecki, 2005; Czarnecki and Eisenecker, 2000) have appeared in the context of computer software development as an innovative paradigm that attempts to make software as general as possible without losing efficiency. GP achieves its goal by identifying interrelated high-level family from a common requirement set. By the application of this technique, especially in design phases, software is developed offering a high degree of abstraction which is applicable to a wide range of situations and domains (Caballé and Xhafa, 2003b).

Moreover, the Model-Driven Development (MDD) (Gomaa, 2005) paradigm and the framework supporting it, namely Model-Driven Architecture (MDA), have been recently attracting a lot of attention given that it allows software developers and organizations to capture every important aspect of a software system through appropriate models (Czarnecki, 2005). MDA (OMG, 2006), in turn, is a novel software development methodology that has proved to work in promoting reusability by providing the necessary tools to fully harness the power of software reuse. The central idea of MDA is to specify a unique and independent view (called Platform Independent Model) of a software artifact and produce several realizations (called Platform Specific Model) automatically via a generator. MDA provides great advantages in terms of complete support to the whole cycle development, cost reduction, software quality, reusability, independence from the technology, integration with existing systems, scalability and robustness, flexible evolution of software and standardization, as it is supported by the Object Management Group (OMG)[13]. MDD and MDA rely strongly on several modeling techniques such as Domain Analysis (Czarbecki, 2000; Gomaa, 2005) as the process of analyzing related software systems in a domain to find their common and variable parts. The usual resulting product is a domain model in the form of a generic architecture, which describes all systems in a domain.

There are many views and opinions about what MDA is and is not. However, the OMG, as the most authoritative view, focuses MDA on a central vision (OMG, 2006; Czarnecki, 2005): allow developers to express applications independently of specific implementation platforms (such as a given programming language or middleware). To this end, OMG proposes the following principles for MDA developments: first, the development of a UML-based Platform Independent Model (PIM), second, one or several models which are Platform Specific Models (PSM). Finally, a certain degree of automation by means of descriptions is necessary for mapping from PIM to PSM (OMG, 2006).

Laforcade et al., 2007, draw an extensive overview from the literature of the promises, challenges and issues encountered when applying principles and theories using MDD/MDA paradigms to model and deploy technology-enhanced e-learning systems. They conclude that this software engineering approach can help designers to reduce the gap between specific instructional requirements and the software architectures that support the implementation, run-time and the regulation of this instruction.

Finally, for the purpose of this thesis, it is crucial mentioning here the Service-Oriented Architecture (SOA) (W3C, 2004) approach, which represents the next step in the software development to help organizations meet their ever more complex set of needs and challenges, especially in distributed systems (Pankatrius and Vossen, 2003). SOA relies on services. According to the W3C Working Group (2004), a service is a set of actions that form a coherent whole from the point of view of service providers and service requesters. In other words, services represent the behavior provided by a provider and used by any requesters based only on the interface contract. SOA benefits software development by mainly dynamically discovering and invoking the appropriate services to perform a request from heterogeneous environments, regardless of the details and differences of these environments. By making the service independent from the context, SOA provides software with important non-functional capabilities for distributed environments (such as scalability, heterogeneity and openness), and makes the integration processes much easier to achieve (Caballé et al., 2007f).

Despite SOA can be realized with other technologies, over the last few years Web-services has come to play the major role in SOA by providing a set of standard protocols that meet the main needs of SOA. The use of widely adopted protocols and standards, such as XML, UDDI, WSDL, SOAP and HTTP (W3C Working Group, 2004), represents the cornerstone of Web-services approach which provides a suitable technology to implement the key requirements of SOA. This is because these protocols allow a service to be platform - and language- independent, dynamically located and

[13] The Object Management Group is found at http://www.omg.com (Web page as of April 2008)

invocated, and interoperable cross over different organization networks. As a result, Web-services technology provides lower costs of integration along with flexibility and simplification of configuration. Next, a short description of the core protocols is provided based on W3C Working Group (2004):

- Simple Object Access Protocol (SOAP): This is a lightweight mechanism used to exchange XML messages between applications regardless of the programming language or operation system. This defines the message's format to be passed between applications, typically (but not exclusively) over the HTTP protocol.

- Web Service Description Language (WSDL): Provides the necessary structure to describe the Web Service's functionalities, by means of providing an XML grammar to be used in the Web Service.

- Universal Description, Discovery and Integration (UDDI): This specification provides the structure used to describe and register a particular process logic, where client applications can find out about services offered, and their locations or addresses.

Therefore, Web-services, based on the eXtensible Markup Language (XML) standard, allow the interoperability of various applications running on heterogeneous platforms, enabling the automation of processes across different applications. A Web-services model abstraction involving the above-mentioned protocols is shown in Figure 1.1. At the base we find the communication protocol (TCP/IP). Transversally to the Web Service stack different protocols can be plugged-in to guarantee good practice regarding security, quality of services and management.

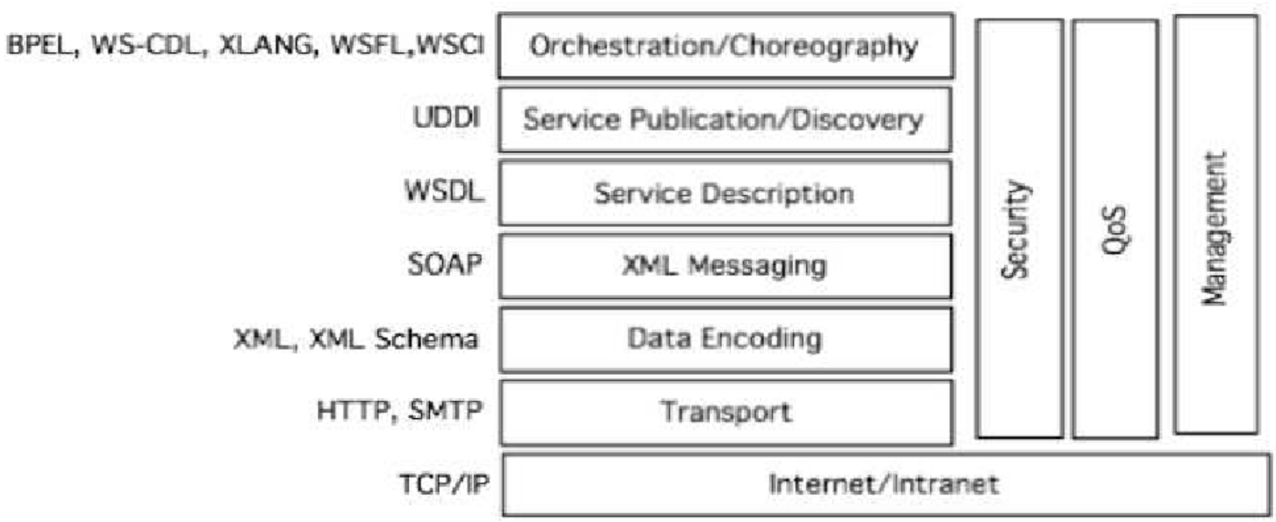

Figure 1.1: Web-services structure.

Even though Web-services have been shown as the ideal technology for implementing SOA, some drawbacks and pitfalls such as performance issues and security incompatibilities to other technologies need to be carefully analyzed and addressed.

Many studies (Endrei et al., 2004; Wang and Fung, 2004) take advantage of the service-oriented approach to improve the development of software systems. One representative application of this approach to the field of e-learning is (Gütl,

2007) who presents an innovative open, distributed service-oriented architecture for flexible teaching and learning activities. In this study, he discusses the experiences gained and shows how a SOA-based system could meet important research and development objectives in this context, such as support for various teaching and learning paradigms, personalization in the retrieval, management and presentation of the relevant information needed to perform the learning activities, and user adaptation to both the system and the learning process by improving the knowledge about the user behavior. However, other equally important benefits from using a service-oriented approach for learning and teaching are not even mentioned in this study, such as reliability, performance, and interoperability with existing and legacy systems.

This thesis takes all these approaches one step further by providing new generation-based techniques that take a higher-level specification of the CSCL domain models and produce its realization via a semi-automatic generator (Czarnecki and Eisenecker, 2000). The users of a generator will see a system that allows them to go from the specification model to a different realization models without having to understand the internal details of the generator (Bell et al., 1994).

1.3 Distributed and Grid computing for enhancing modern needs in CSCL

According to Foster et al. (2001) and Foster and Kesselman (1998), Grid computing has emerged as a way of capturing the vision of a networked computing system that provides broad access not only to massive information resources, but to massive computational resources as well. The concept of computational Grid has its origins in wide-area distributed computing, and extends to a large-scale, flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources.

Grid architecture (Foster et al., 2001) is found in the form of five layers, which may be distributed in different levels (Figure 1.2): Fabric at the resource level, Connectivity, Resource, and Collective, at the core Grid level, and Application, at the user level. Detailed information of each layer can be found in the literature (Foster et al., 2001).

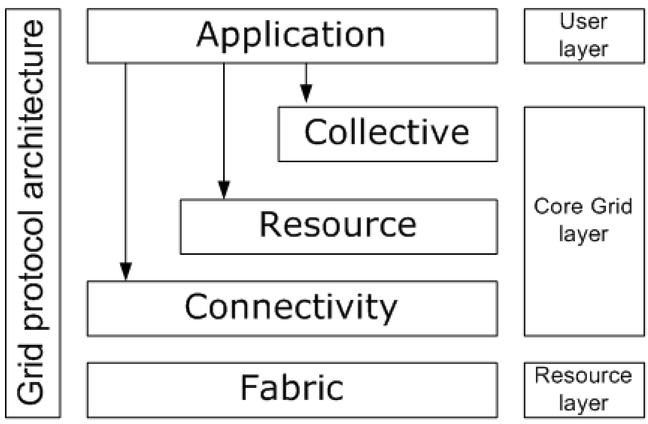

Figure 1.2: The layered Grid architecture.

A fairly amount of e-Learning Grids have appeared over the last years, which leverage Grid architectures to yield more powerful resources for educational organizations (Pankatrius and Vossen, 2003; GuiLing et al., 2005b). Representative e-Learning Grids are OntoEdu (Guang-zuo et al., 2004) SELF (Abbas et al., 2005), and CoAkTing (Shum, 2002), which overcome important non-functional requirements arisen in this context, such as scalability, availability, and distribution of computing power as well as storage capability (Caballé et al., 2004).

From this approach, Grid provides an ideal context for supporting and producing major benefits for CSCL applications. Such important features include (Foster et al., 2001; GuiLing et al., 2005b): large scale of Grid infrastructures, wide geographical distribution of resources, multiple administrations from different organizations, transparent and dependable access as well as the capability of granting access to shared, heterogeneous resources in very dynamic environments. Considering these benefits provided by Grid it is possible for educational organizations to make use of true collaborative learning environments that enable the involvement of large number of single/group participants, who can potentially belong to many different organizations, possibly situated at very different locations, and transparently share a huge variety of both software and hardware resources while enhancing human-to-human interactions (e.g. through a friendly 3D-based user interface) (Caballé et al., 2004).

Therefore, leveraging the inherent performance potential of Grid infrastructure for CSCL applications makes it possible to greatly enhance the collaboration experience. Moreover, the combination with other technologies, such as Service-Oriented Architectures and Web-services[14], allows developers to cope with essential issues in CSCL, such as integration, interoperability, fault-tolerance (reliability) and flexibility so as to meet the needs of different, heterogeneous and legacy environments (GuiLing et al., 2005b). In this context, it seems clear from the literature the key role played by SOA-based architectures, and in particular Web-service paradigm. Web-service technologies provide both interoperability to overcome the great complexity of Grid middleware and ease for the management and delivering of heterogeneous, complex learning content and courses (GuiLing et al., 2005a).

Nevertheless, there is still a strong lack of support for specific situations and sub-domains of e-Learning such as CSCL. Just a few Grid systems are intended to offer specifically CSCL support (Bote et al., 2007), and it is demanded more penetration of Grid in this context (Gleeson and Pahl, 2007) given that its inherent potential, in terms of sharing huge amounts of learning data and resources during the collaboration at reasonable cost.

Despite that the research areas previously reviewed have been largely investigated, their analysis revealed that they are still far from being mature and a lot of effort is still needed to overcome considerable barriers that form part of their very rationale. To this end, this thesis' results and experiences are expected to contribute to the respective research fields and, in particular, the research community of the UOC, by means of the provision of distributed CSCL systems that will hopefully fill an important gap existing in current research. Next chapters provide a detailed description of these entire contributions.

[14]The Web Services Architecture Document is found at http://www.w3.org/TR/ws-arch/ (Web page as of April 2008)

Chapter 2

A conceptual framework for interaction data management and knowledge extraction

The success of CSCL applications depends to a great extent on the capability of such applications to embed information and knowledge of group activity and use it to achieve a more effective group monitoring (Caballé et al., 2006). A large amount of information data is generated from long-term collaborative interaction which includes complex issues of the collaborative work and learning process (e.g., group well-being (McGrath, 1991) as well as self, peer and group activity evaluation). Some of this information may be produced from specific data sources such as ad hoc questionnaires and, due to its high degree of informality, needs to be processed and analyzed manually. Consequently, the efficient embedding of all this information and of the extracted knowledge into CSCL applications sets the basis for enhancing support, awareness and feedback to achieve a successful learning process in collaborative environments.

In this thesis, it is considered that asynchronous collaboration generates quantitative information in the form of events as a result of the users' interaction with the system's resources and other users. Quantitative information generated from synchronous collaboration can be managed by applying a pre-structuring process where users' interactions are labeled with certain indicators according to a rhetorical exchange structure (Daradoumis, 1995) that models the various types of interactions at different levels. All this information can be easily collected and automatically processed and analyzed by computers as a quantitative data source. The knowledge extracted by this process can then be used to facilitate a continuous monitoring of the learning activity, providing group members with appropriate support, as well as awareness about what is happening during collaboration. Furthermore, the constant and fast processing (Paniagua et al., 2005) of the quantitative data as well as their systematic analysis based on principled indicators that measure the type and the degree of group members' participation, may positively impact on participant's motivation, emotional state and problem-solving abilities and as a result enhance on-line collaborative learning (Caballé et al., 2005b).

Furthermore, qualitative information is collected from ad hoc questionnaires which are regularly filled out by group members, reporting human and behavioral aspects of collaboration as well as evaluating the collaborative learning experience. Participants qualify their own emotional and motivational state within the learning group as well as evaluate the participation and learning activities of their peers. The aim of this approach is to provide both a deeper understanding of collaboration and a more objective assessment of individual and group activity.

The ultimate aim of this research work is to extract relevant knowledge of the collaboration process from all possible sources. Note that in this context information refers to quantitative and qualitative data generated by the learning group whereas knowledge refers to the result of the treatment of this information through analysis techniques and interpretation. This knowledge will be fed back and presented to the learning group and its tutor for awareness and scaffolding purposes.

The management of both quantitative and qualitative information generated in both synchronous and asynchronous collaboration aims at achieving three main goals: (i) provide an analysis of the group's performance at three levels, namely collaborative learning outcome, group functioning and scaffolding (Caballé et al., 2007f), by obtaining and classifying the necessary information gathered from the collaborative activity into these three essential categories; (ii), implement an effective way to collect, analyze and present this information given that the large amount of information generated during online group activity may need much time to be processed; (iii) embed the information and knowledge obtained into CSCL applications efficiently so as to facilitate both tutors to monitor the learning activity and group members to get as much and effective awareness and feedback as possible.

In order to achieve these goals, a conceptual model is first proposed for data analysis and management that identifies and classifies the many kinds of indicators (variables) that describe collaboration and learning into high-level potential aspects of collaboration. Then, a process is described that, at a first step, collects and classifies both the event information generated asynchronously from the users' actions and the labeled dialogues from the synchronous collaboration according to these indicators. For efficiency purposes, this information may then be structured further in a way that facilitates its faster processing and analysis (Caballé et al., 2005a). The last stage of this process consists of interpreting the analysis outcomes and communicating the knowledge extracted to the group members for awareness and feedback purposes as well as to the tutors to track the collaborative learning process more effectively (Caballé et al., 2008a).

2.1 A conceptual model for data analysis and management

In the context of both asynchronous and synchronous, collaborative learning practice, a conceptual model for data analysis and management is intended to model different aspects of interaction and thus at helping all the actors involved understand the outcomes of the collaborative process classified into three generic group activity parameters: the members' contributing behavior to the task (the product of collaboration), the functioning of the group (the interaction processes underlying the collaborative learning activities, such as participation behavior, role playing, etc.), and individual and group scaffolding (social support and task- or group functioning-oriented help).

Indeed, the specification of high-level collaborative learning processes constitutes the first step toward the classification of the many different variables that characterize collaborative interaction as well as the identification and measurement of these variables in terms of the user and system specific events (or actions). This conceptualization enables the construction a computational model to gather information in a structured manner and consequently to provide an easier and more efficient further processing and analysis of this information through different techniques (such as statistical and data mining, social network analysis etc.).

The conceptual model proposed in this thesis starts up from this integrated approach, in particular, from the mentioned high-level indicators of the quantitative dimension. This approach is then extended with new and finer-grained processes and indicators in a top-down fashion, from high-level processes to specific indicators. This way, each group members' action and contribution as well as its impact on the group activity can be precisely identified, classified and assessed according to real skills and intentions shown by the participants. The aim is to interpret the analysis results and extract, reveal and provide the actors with valuable knowledge for each of the three high-level collaborative learning processes.

The construction of this conceptual model is based on knowledge acquired primarily by an in-depth statistical analysis of a large number of log files describing group interaction from a variety of real collaborative learning experiences, and secondarily by a qualitative analysis performed by the tutors who conducted the experiences. The ultimate aim of this study is two-fold: First, to show to what extent the gathering, pre-processing, and analysis of collaborative data can be automated. Second, to infer how the knowledge extracted from the analysis can be used as feedback to report and predict behavior as well as to detect problematic situations.

This research takes place in the context of online learning groups composed of students from several distance learning undergraduate courses at the Open University of Catalonia (UOC) being set the task of working on real, long-term, complex, collaborative problem-solving situations. As such, it is important to present first a sufficient description of the workings of the case study that was used for the purpose of the analysis on interaction. The results of this analysis defines and obtains several indicators of group activity as regards to each of the three high-level collaborative learning processes; a general description for each indicator is provided below. Note that the whole learning activity at the UOC occurs largely in asynchronous mode rather than synchronously. Thus, for the time being, this research is mainly focusing on asynchronous interaction, even though synchronous mode is also considered when needed.

2.1.1 Case study description

This study is based on real collaborative learning experiences carried out in the scope of online distance learning undergraduate interdisciplinary courses, such as *Application of Information Systems to Business*, and *Software Development Techniques* at the UOC. The first experience ran over a period of 14 weeks and involved 2 tutors and 122 students distributed into 21 online groups of 5 to 6 members. Students had to collaborate and work out a case study that simulated a real project in a business or organization. The second experience lasted the same period of time, involved 2 tutors and 60 students distributed into 12 online groups of 5 members, and was based on the Project-Based Collaborative Learning paradigm.

The implementation of each collaborative learning practice consists of five well-differentiated and structured sub-problems (phases): problem specification and planning, design, implementation, testing, and documentation and product delivery. In each phase the tutor should assess both individual and group contribution, so even though students work together to achieve a learning goal and submit a common product, they are evaluated by means of their particular contributions as regards the product and the collaborative process itself. Individual assessment also depends on the specific role a student plays at a particular phase. Roles are switched among group members while passing from one phase to another, so a student plays a different role in each phase (for more details see Daradoumis, el al. (2006)).

The whole project was carried out mostly asynchronously; synchronous interaction occurred in few specific cases of decision-making. All asynchronous collaborative interactions took place on the Basic Support for Cooperative Work (BSCW) system, a groupware tool that enables asynchronous and synchronous collaboration over the web (Bentley et al., 1995). BSCW offers shared workspaces that groups can use to store, manage, jointly edit and share documents, realize threaded discussions, etc.

To structure the whole collaborative learning process, two particularized shared workspaces were set in the BSCW system. The first one is a general workspace, which could be accessed by all students of the online class. The main purpose of this workspace was to let students interact with each other in order to form the online learning groups. In addition, it was used to effectuate specific debates, which form part of the project requirements and involve all students, as well as to share important information about the project among tutors and students. The other workspace type was a private space designated to house each online group, that is to record and structure the interaction of its members that aims to achieve the project target goals through the resolution of the specific tasks and problems the project consists of.

The successful realization of these experiences (as well as of others that were carried out in subsequent semesters) provided a large volume of interaction data that constituted a valuable source for the analysis. This information was maintained in the form of event log data and was generated automatically by the BSCW by registering the information related to different types of actions done by the users of the applications.

This approach first builds a conceptual model of analysis of interactions which relies on the theoretical principles and indicators of effective collaboration. This provides a principled and effective manner to classify the information generated from group interaction; it also facilitates the processing and analysis of this information and knowledge extraction. Next, both the event information and the users actions that generates events are explored further.

2.1.2 User events

User events in the form of generic indicators are defined to describe the three main high-level collaborative learning processes, which take place in group educational activities: Collaborative learning product, group functioning, and scaffolding (see Figure 2.1). In order to describe each of these general categories, similar terminology to the one used in the Basic Support for Cooperative Work (BSCW) system is employed to refer to the actions that can be carried out in an asynchronous groupware platform. However, they are general enough to be abstracted and represent all the typical and basic actions encountered in any asynchronous groupware platform. In addition, the terminology used for labeling the

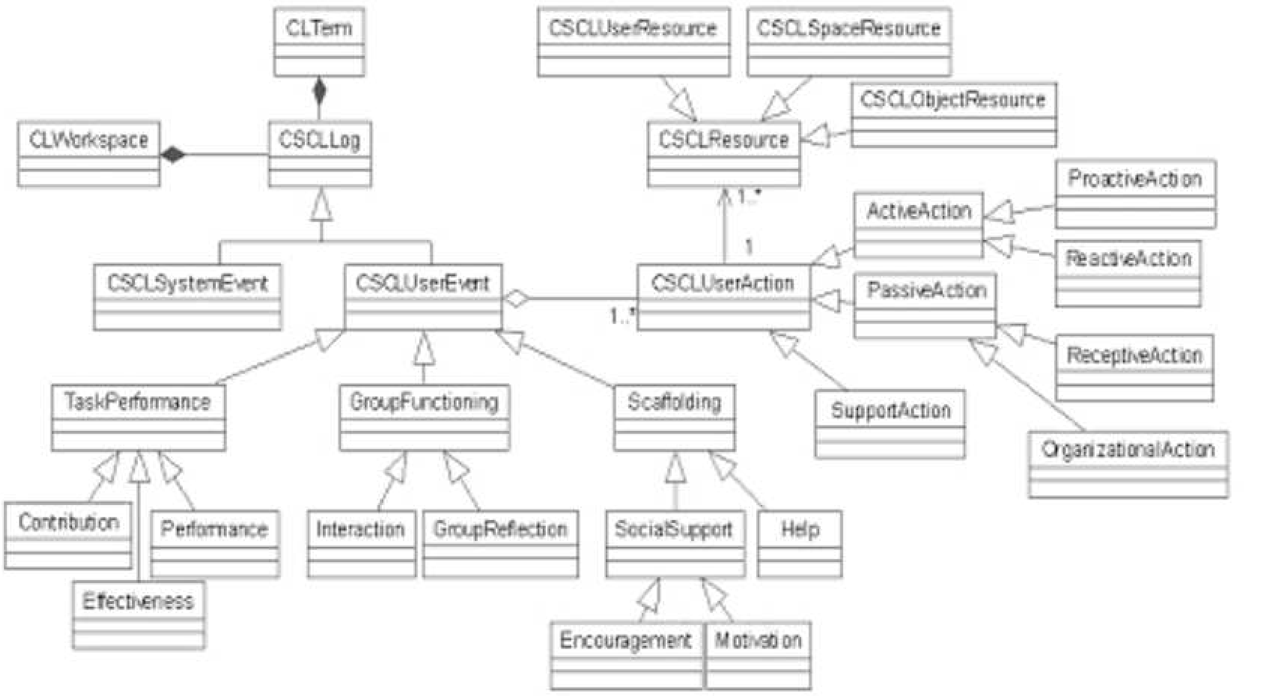

Figure 2.1: A hierarchy to collect and classify all events generated during the group activity.

dialogues generated in collaborative synchronous environments is based on the rhetorical exchange structure explained in Daradoumis (1997).

2.1.2.1 Collaborative learning product

This is the first top-level activity parameter featuring the production function and task performance of on-line groups. It is characterized by the type of actions (events) that capture and describe the functional knowledge, cognitive processes and skills of the students and the group as a whole in solving problems and producing learning outcomes in a collaborative learning practice. It is used to analyze and evaluate the individual and group effectiveness as far as task achievement concerns. It can be measured as a qualitative and quantitative parameter by the type of user task-based actions that represent contributions which express basic and supporting active learning skills as well as perception skills. Table 2.1 shows the mid- and low-level indicators in the form of the skills and sub-skills that should characterize the students who participate in a collaborative situation in order to achieve effective group and individual performance of the task and thus obtain a successful learning outcome. In measuring each indicator (or skill), it is associated with both the actions that students perform in an asynchronous (A) environment and the type of dialogues carried out synchronously (S). A similar terminology is employed to the one used in the BSCW system to refer to the actions that can be carried out in any groupware platform. Indeed, they are general enough to represent all the typical and basic actions encountered in every groupware platform.

Table 2.1: Indicators (skills) that model *task performance*

Skills	**Sub-skills Learning outcome contribution**	**Asynchronous actions (A)** **Synchronous communicative acts (S)**
Basic active learning skills	Information generation	Create doc/note (A) Describe / Edit doc (A) Adjust (S)
	Information elaboration	Version/Replace doc (A) Elaborate (S)
	Information revision	Revise/Branch doc (A) Revise (S)
	Information reinforcement	Create-Noteboard doc/URL /Notes (as an attachment) (A) Extend (S)
Information processing (perception) skills	Information acknowledgment	Read event (A) Give consensus (S)

2.1.2.2 Group functioning

This is the second top-level activity parameter which is made up of the type of events that represent and are used to measure and analyze the individual and group effectiveness regarding participation and interaction behavior that facilitate the group's well-being function (MacGrath, 1991). As a quantitative parameter, it enables the measurement of important participant contributions (in terms of specific types of user actions) which indicate skills related to: active or passive participation, well-balanced contributions and role playing, participation quality and communication flow among group members, as well as the necessary skills that facilitate and enhance group interaction, namely active processing skills (such as task, workspace and communication processing skills). In addition, interaction behavior can also be measured as a qualitative parameter by group reflection (i.e. group and individual self-evaluation). Table 2.2 shows the mid- and low-level indicators in the form of the skills and sub-skills that students should exhibit in order to enhance participation, promote better communication and coordination, and thus achieve the effective interaction and functioning of the group in a collaborative situation. Again, to measure each indicator (or skill), it is associated with specific student actions which best describe each skill to be accomplished.

2.1.2.3 Scaffolding

This last top-level activity parameter is specified by the type of events that refer to social support among members as well as to task- or group functioning-oriented help provided to a participant who is not quite able or ready to achieve a task on his or her own. As for the former, the event information is considered that includes actions which support and promote group cohesion, such as motivational and emotional support, conflict resolution, etc. As for the latter, it is focused on those specific actions designated to provide effective help to the peers when they need it during the collaborative learning activities. The participants' actions aiming at getting or providing help are classified and measured according to whether they refer to the task or group functioning.

Table 2.3 shows the different types of social support and help services (Webb, 1992) that have been identified and accounted for in the proposed model. Scaffolding can be provided explicitly if a member asks for it, or implicitly when a specific need or problem is detected. The latter is achieved through awareness and feedback. More specifically, once the group cohesion, task performance, and group functioning parameters have been measured and analyzed, the group participants (including the mediator) or the system itself (by means of an intelligent agent) can be aware of what is happening during the group activity. Thereafter, they can decide to provide the participants an adequate scaffold associated to social support (motivation, encouragement, conflict resolution, etc.) or supply them with a specific help service related to the task itself or group functioning (e.g. member participation).

2.1.3 User actions

In a collaborative learning experience, the group activity is driven by participants' actions on the generic collaborative learning resources and these actions are aggregated to the user events to form another hierarchical tree (included in Figure

Table 2.2: Indicators (skills) that model *group functioning*

Skills	**Sub-skills (Group functioning contribution)**	**Asynchronous actions (A)** **Synchronous communicative acts (S)**
Active participation behavior and peer involvement skills	Participation in managing (generating, expanding and processing) information	Create Event, Change Event, Read Event (A) Take-initiative, Provide-info, Share-info (S) Request/Suggest-action, Listen (S)
Social grounding skills	Well-balanced contributions, adequate reaction attitudes, and role playing	Version/Replace doc (A) Create Event, Change Event, Read Event, Move Event (A) Provide-acknowledgment/answer/solution (S) Assess, Give/Take-turn, Perfom-role (S)
Task processing skills	Task planning/distribution	Create/Link Appointment (A) Create/ChangeAccess WSCalendar (A) Coordinate-task, Plan, Distribute-time (S)
	Task (and knowledge) management Work load distribution	Create Folder (A) Create Notes (as a contribution in a bulletin board) (A) Build-workspace, Distribute-workload (S)
Workspace processing skills	Workspace organisation and maintenance	Move event (cut, drop, copy, delete, forget) (A) Organize, Order, Clear-out (S)
Communication processing skills	Clarification	Change Description / Change Event doc (A) Change Description url (A) Clarify (S)
	Evaluation	Rate document/url (A) Evaluate (S)
	Description (illustration)	Edit/Change Description Folder (A) Change Description Notes (A) Illustrate (S)
	Communication Improvement	Edit Note ; Chvinfo/Chvno/Checkin/Checkout doc (A) Rename Folder/Notes/doc/url/ (A) Rephrase, Reformulate (S)
	Meeting accommodation	ChangeDesc/ChangeDate/ChangeLocation (A) Appointment (A) Arrange, Accommodate (S)

Table 2.3: Indicators that model *scaffolding* (for both asynchronous and synchronous collaboration)

Social support
Members’ commitment toward collaboration, joint learning and accomplishment of the common group goal
Level of peer involvement and their influential contribution to the involvement of the others
Members’ contribution to the achievement of mutual trust
Members’ motivational and emotional support to their peers
Participation and contribution to conflict resolution
Help services
Help is timely
Help is relevant to the student’s needs
Help is qualitative
Help is understood by the student
Help can readily be applied by the student

Table 2.4: Excerpt of a generic coding scheme for asynchronous environments.

Code	Action	Event type	Skills	Category
cdc	Create document	Creation	Information generation	Contribution
cda	Create document	Activity	Active participation	Interaction
cdr	Create document	Reply	Information revision	Effectiveness
cde	Create document	Evaluation	Task contribution	Group reflection
cnc	Create note	Creation	Information reinforcement	Performance
cns	Create note	Support	Members’ involvement	Motivation
cnr	Create note	Reply	Information revision	Effectiveness
cna	Create note	Activity	Active participation	Interaction
rdp	Read document	Processing	Information knowledge	Performance
rda	Read document	Activity	Passive participation	Interaction
mdr	Modify document	Revision	Information revision	Contribution
mda	Modify document	Activity	Active participation	Interaction
mde	Modify document	Evaluation	Task contribution	Group reflection
rde	Replace document	Elaboration	Information elaboration	Effectiveness
rdc	Replace document	Activity	Active participation	Interaction
rdv	Replace document	Evaluation	Task contribution	Group reflection

2.1). In this hierarchy, at a first level, the difference between *active* and *passive* user actions is considered depending on whether or not the student contributes directly to achieving the group objective. At this same level, the *supportive* action (i.e. help, motivation and encouragement) is also considered and constitutes another distinct category. Further, an active action is particularized in proactive (i.e., the user takes the initiative) and reactive (i.e., the user replies) while passive actions distinguishes between receptive (i.e., the user reads others’ contributions) and organizational (i.e., coordination) actions. Note that *CLWorkspace* in Figure 2.1 refers to the log file aggregating event information that is generated in a given workspace. Such a workspace may correspond to a whole group or to a phase within a group activity.

In order to correctly classify the user actions on the resources during group activity according to the event hierarchy, a classification process is proposed consisting in a coding scheme (see Table 2.4) for asynchronous environments based on the conceptual model proposed. By means of this coding scheme, each participants’ action is interpreted depending on the type of event that was involved, such as in response to a previous contribution. This represents the essential information in the identification of the real intentions or skills shown by the user (e.g. creating a note during a debate can be interpreted as either revision or reinforcement of the information depending on whether the note was created in the context of a reply, an observation, agreement, etc.). Then, the user events are uniquely codified according to both the user action performed and the real user skill identified in the context of the action. Thus, for instance, creating a replying note is codified with a unique code. Finally, the user event is categorized into one of the three above-mentioned group activity indicators (see Table 2.1 through Table 2.3) according to this coding scheme. For more information on this classification process, see Section 2.3 and Figure 2.2.

Given that this classification process is highly generic, only the most abstract form of categorization is provided based on the above-mentioned generic event hierarchy (Figure 2.1). Thus, the specific applications using this process should categorize their event information according to their particularization of this categorization (e.g., the discussion process that is analyzed in the next section). Note that although it is possible to use the same classification process for both synchronous and asynchronous environments, this research will focus on the latter as this is still the most usual way to collaborate in on-line collaborative learning environments and permits the complete automation of the classification process. In contrast, in synchronous environments most of this process has to be performed manually and it needs a different coding scheme to codify the user actions.

2.1.4 Qualitative information

Finally, qualitative information about *group functioning* and *scaffolding* is also extracted by specifically designed structured and non structured questionnaires which are filled by group members at the end of each collaborative problem-solving phase. Structured questionnaires provide a predefined set of answers to choose and as a result can be collected and processed by computers whereas non structured questionnaires present a high degree of informality and thus need to be processed and interpreted manually. Table 2.5 shows a generic questionnaire scheme which is eventually elaborated and adapted to the particular problem-solving situation.

2.2 Modeling interaction in the discussion process

From the general indicators seen, this section examines how precisely learning and knowledge building can be supported in the specific case of an asynchronous collaborative discussion in a virtual learning environment (Caballé et al., 2008a). To this end, a conceptual sociolinguistic framework is defined for modeling dialogue and understanding how learning evolves and how knowledge is constructed during the discussion process. This framework results from the particularization of the general three indicators described in the last section. This way, the specific interaction types identified here correspond with the particular case of a collaborative discussion process, which is described next and represents the empirical context of this thesis, serving to validate the general model proposed.

2.2.1 Learn by discussion

The discussion process plays an important social task in CSCL where participants can think about the activity being performed, collaborate with each other through the exchange of ideas that may arise, propose new resolution mechanisms, as well as justify and refine their own contributions and thus acquire new knowledge (Salomon, 1993). Indeed, learning by discussion when applied to collaborative learning scenarios can provide significant benefits for students in the context of project-based collaboration learning, and in education in general (Stahl, 2006).

Moreover, learn by discussion in the context of CSCL fits the current shifting from a traditional educational paradigm (i.e., centered on the figure of a masterful instructor) to an emergent educational paradigm (Juan et al., 2008) which considers students as active and central actors in their learning process while the instructor's role is moving from one related to a knowledge transmission agent to another related to a specialist agent who designs the course, guides, assists and supervises the student's learning process (Simonson et al., 2003).

Table 2.5: A generic questionnaire scheme for extracting qualitative information about *group functioning* and *scaffolding*

Actions carried out to plan, manage and make the group activity evolve (Text)
Actions carried out to organize and maintain the group workspace (Text)
Actions carried out to coordinate the group effectively (Text)
Description of the most relevant conflicts encountered in the group and the way they were resolved (Text)
Assessment of own participation in the learning group (0 - 5)
Assessment of the level of engagement of the other group members (0 - 5)
Description of the problems that affected group dynamics in terms of engagement, communication, organization, and so on (Text)

To this end, a complete discussion and reasoning process is proposed based on three types of generic contributions, namely specification, elaboration and consensus. Specification occurs during the initial stage of the process carried out by the tutor or the group coordinator who contributes by defining the group activity and its objectives (i.e. statement of the problem) and the way to structure it in sub activities. Elaboration refers to the participant contributions (mostly students) in which a proposal, idea or plan to get to the solution is brought up. The other participants can elaborate to this proposal according to the different types of communicative acts the proposal can be structured, such as questions, comments, explanations and agree/disagree statements. Finally, when a correct proposal of solution is achieved, the consensus contributions take part to approve it (this includes different consensus models such as voting); when a solution is accepted the discussion closes up (Caballé et al., 2004).

In the real context of this thesis, which is the virtual learning environment of the Open University of Catalonia (UOC)[1], most part of the courses' curricula includes the participation of students in on-line discussions with the aim of sharing and discussing their ideas and as a result acquiring more knowledge. Given the added value of asynchronous discussion groups and the extensive use of online debates, as one of the main elements of the UOC's pedagogical model, it is essential to provide adequate on-line tools to support the whole discussion process, which also includes students' monitoring and evaluation.

Next, a complete and concrete interaction data analysis model centered in the discussion process is described. Note that this model is a subset of the general model described in the previous subsection. The ultimate aim is to validate the general model by providing theoretical and empirical work on a concrete and practical aspect of interaction data analysis and management.

2.2.2 A model for managing a discussion process

One important issue to consider in the context of a discussion process is the types of interaction that occur and subsequently the knowledge which is manifested in an asynchronous collaborative discussion. This approach aims at identifying the various types of interaction produced and examining how an interaction type is related to the learning that results from it. As a result, this framework allows the study of how knowledge is transformed and becomes common to all discussion members (Caballé, at el., 2008a).

In particular, this section examines how the building and distribution of knowledge is manifested in the context of student-student interaction and how it can be studied in a virtual learning environment. This involves the definition of appropriate collaborative learning situations and the distinction of two levels of student interaction, the discourse and the action level.

At the discourse level, the essential element is the interaction among peers (participants need to interact with each other to plan an activity, distribute tasks, explain, clarify, give information and opinions, elicit information, evaluate and contribute to the resolution of problematic issues, and so on). At the action level, task objects (e.g., documents, graphics)

[1]the UOC is found at http://www.uoc.edu. The UOC offers full distance education through the Internet to more than 45,000 students

are created and manipulated. This approach focuses more at the analysis of the discourse level by seeing discourse as a medium and means through which the building and distribution of cognition is effected.

To satisfy course assessment requirements, discourse contributions also need to be evaluated as effectively as possible. Evaluation of hundreds of contributions in a multi-member discussion can be a tedious task for tutors and should be adequately supported. Moreover, self and peer assessment should be also encouraged and facilitated by intuitive means. Then, a dialogue model of asynchronous discourse is to be provided, which is capable of capturing, analyzing and evaluating both the process and the result of the building and distribution of knowledge. This model should be mainly defined in terms of types and structure of student-student interaction.

In particular, the framework proposed in this paper to support this model is based on an integration of several models and methods: the Negotiation Linguistic Exchange Model (Martin, 1992); a model of Discourse Contributions (Clark and Schaefer, 1989); and, the types of learning actions underlying a participant turn (Self, 1994). The structure of a long interaction is constructed cooperatively by using the exchange as the basic unit for communicating knowledge. Following Martin (1992), three general exchange structure categories are considered: give-information exchange, elicit-information exchange and raise-an-issue exchange, which consist of different types of moves (Schwartz, 1999) and describe a generic discourse goal. More specifically, the goal of the actor who initiates the give-information exchange is to inform his/her partners about a certain situation with the aim to change the partners' mental states. Informing includes moves that explain, give an opinion, describe or remind a situation in different ways. The actor goal of the second exchange is to elicit the partners' state of mind (knowledge, beliefs, attitude, desire or abilities) of a situation which the actor is not aware or certain about. The actor goal of the third exchange is to raise an issue (a problem or question) to be resolved by the participants, which causes to explore their state of mind (knowledge, beliefs, etc.).

According to Martin (1992), there is a move that constitutes the "obligatory move" of the exchange, since it either carries or indicates completion of the discourse goal for which the exchange is initiated. The obligatory move of each of the above exchanges is: the first move of the give-information exchange, the second move of the elicit-information exchange and the third move of the ascertain-information exchange.

According to Clark and Schaefer (1989), each move is seen as a contribution to discourse. This means that in a cooperative conversation, contributions are regarded as collective acts performed by the participants working together, resulting in units of conversation - typically turns (moves) - that aim to make a success of the discourse they compose. Yet, not all moves contribute in the same way toward the successful completion of the exchange.

Some moves have a pure contributing function toward the realization of the obligatory move of the exchange. This is the case of the first move of the elicit-information exchange, as well as of the first and the second moves of the ascertain-information exchange. In fact, without the presence of those moves, the obligatory move cannot be realized; thus, those moves really contribute toward the realization of the obligatory move. Consequently, it is stated that successful realization of the obligatory move conveys evidence of (initial) success of the exchange (Clark and Schaefer, 1989). In contrast, the other moves have a rather supporting function (provide evidence of support) toward the definite completion of the obligatory move and consequently of the exchange. This is the case of the follow-up moves of the three exchanges. Supporting moves are optional, so they may not be realized. In such a case, they convey an implicit support toward the obligatory move, that is, toward the definitive completion of the exchange.

Based on the work of Self (1994), Pilkington (1999), and Soller (2001), partners are involved in a process of realizing a number of learning actions which lead to the completion of the exchange goal. Each move type captures and controls the evolution of the learning action performed by a participant by setting the expectations of the type of learning actions which has to be realized next by the other participants so that the goal set by the initial move be accomplished.

Both the quantity and the quality of the several move types performed are measured by the collaborative effort of the members involved to achieve the discourse goal of an exchange. The term collaborative effort means both the number of contributing and supporting moves issued by a participant, which indicates an active participation (distinguishing between proactive and reactive one) or passive one, and the type and effectiveness of these moves (which indicates the way a participant contributes toward the achievement of the shared discourse goal, as regards knowledge possession and transfer, reasoning capability and positive attitude). The tutor measures move effectiveness by assessing the quality of their content. In addition, peer assessment can be effected to complete the evaluation of each contribution made. The role these moves plays in the exchange as well as the degree of success of that role determine the successful completion of the exchange goal.

Completion of an exchange expresses the mutual beliefs of all participants about the accomplishment of its discourse goal. Moreover, it implies the achievement of a certain degree of knowledge building and distribution among the different participants. This degree can be deduced and measured by exploring the principal interaction indicators proposed by this model. For each participant the model measures: the total number of moves created, his/her participation behavior (proactive, reactive, supportive, or passive), the effectiveness and impact that each move has in the discourse and in the achievement of the current discourse goal, as well as the evaluation of the move content and significance by his/her peers and the tutor.

In general, the three types of exchanges represent standard discourse structures for handling information and suggest a certain type of knowledge building, as a result of giving and eliciting information or working out a solution on an issue set up. These discursive structures enable the participants to take turns, share information, exchange views, monitor the work done and plan ahead. Most importantly, they provide a means to represent and operationalize the cognitive product at individual level, that is, the way the reasoning process is distributed over the participants as it is shared in a collaborative discourse.

Consequently, interaction analysis takes into account both the way the interaction is structured and the types of contributions which are explicitly defined and expressed (see Table 2.6). The analysis of these interactions yields very useful conclusions on aspects such as individual and group working, dynamics, performance and success, which allows the tutor to obtain a global account of the progress of the individual and group work and thus to identify possible conflicts and monitor the whole learning process much better.

Table 2.6: List of the exchange moves and categories to classify a contribution.

Moves	**Categories**
support	Greeting Encouragement Motivation
request	REQUEST-Information REQUEST-Elaboration REQUEST-Clarification REQUEST-Justification REQUEST-Opinion REQUEST-Illustration
inform	INFORM-Extend INFORM-Lead INFORM-Suggest INFORM-Elaboration INFORM-Explain/Clarification INFORM-Justify INFORM-State INFORM-Agree INFORM-Disagree
set-up-an-issue	PROBLEM-Statement
provide-solution	PROBLEM-Solution
consent-solution	PROBLEM-Extend solution PROBLEM-Assent solution

A further innovation of this model is that it allows participants to end up an exchange which took several moves to conclude by ”replaying” the main contributing move of the exchange. For instance, in a set-up-an-issue exchange, a solution move may not be sufficiently complete and thus has to be further elaborated, corrected or extended. To that end, another participant has the option to provide an amplify-solution move which completes the initial solution. In general, a *replay* move can be used to resume all the changes produced from the initial appearance of an exchange goal to be achieved to its final conclusion and acceptance by all participants. This can be useful both to reinforce the fact that the goal of the exchange has been completed successfully and to explicitly indicate the progress achieved in the participants' process of knowledge building (especially as regards the participant who provided the main contributing move of the exchange). A complete set of categories or types of contributions and the context of moves where they are found is proposed in Table 2.6

Finally, the system requires the participant to commit certain action to indicate s/he has read a certain contribution, such as send a reply and assent the contribution. The aim is both to provide reliable indicators on the number of contributions read and to promote the discussion's dynamics by increasing the users' interaction with the system.

2.2.3 Indicators used to assess participation behavior, knowledge building and performance

Based on the categories of contributions identified in Table 2.6, the definition and measurement of the indicators used to assess participation behavior, knowledge building and performance are discussed here.

- Participation behavior indicators are distinguished into *proactive*, *reactive* and *supportive* (or *assentive*). Participants are proactive when they take the initiative to open a new exchange of the type *give-information*, or *raise-an-issue*. Participants are reactive when they reply to moves such as *elicit-information*, *set-up-an issue/problem*, or *provide-solution.* Participants are supportive if they give their assent to previous contributions. In that case, a supporting value is defined which is assigned a default numerical value 1 which means that the move fully supports and recognizes the value, contribution and effectiveness of a previous move it refers to. If several supporting moves refer to a particular move M, it implies a broader consensus about the impact of M, which increases M's impact value to 1.

- *Passive* participants are considered those who just read others' contributions, as well as the ones who also evaluate the usefulness of these contributions. Passivity becomes an essential indicator for the discussion process' dynamics as it identifies certain important profiles of the participant, such as arrogance (participant who just contributes but does not read the contributions of others) and also promotes reactive attitudes and social grounding skills (Daradoumis et al., 2006) by engaging the participant in the collaborative process.

- The *impact* value is assigned an initial (default) numerical value between 0 and 1 which is modified (increased or decreased) according to the impact (number of reactions received) that the move M has on the dialogue and on the achievement of the current discourse goal and task. If the reaction is positive (the move M is being assented), then M receives a positive one (+1) point. If the reaction is negative (M is not assented) then it receives a negative 0.5 points. The points received by a reaction move depends on the type of learning action underlying the move and take on the default value of the move's impact value. The final value is obtained by the mean value of all moves involved in move M.

- The *effectiveness* value of a move is calculated by the mean value of the number of assent moves received. An assent move M is identified and recorded after a participant receives M and consents it. Note that only give-information and raise-an-issue exchange acts can be assented. A negative assent requires a reply move on M to provide further information to reason why M has not been assented, which generates another move in the current discourse.

- Tutor and peer *assessment* indicators are to evaluate both the quality of the contribution's content by the lecturer monitoring the discussion process and the usefulness of the contribution by the student participating in the discussion. Both indicators are on the scale 0-10 so as to be accurate in providing mean values of them.

All these quantitative and qualitative indicators are to be weighted adequately according to the specific goals and procedures of each discussion. To that end, a fully customizable environment is necessary to parameterize and adjust each indicator with an appropriate weight by the tutor at any moment of the discussion process.

So far it has been shown how the information generated in collaborative learning activities can be captured, structured and classified at several descriptive levels. This fact can significantly improve the way a groupware system used for learning and instruction can collect all the necessary information produced from the user-user and user-system interaction in an efficient manner. The next problem is how to analyze the information stored and to consider what kind of knowledge should be extracted to be fed back to the participants in order to provide the best possible support and monitoring of their learning and instructional processes; that is, extracted knowledge that is relevant and applicable to the needs both of students and tutors. Finally, there is a need to provide an efficient and robust computational approach that enables the embedding of the collected information and the extracted knowledge into a CSCL application. Next section outlines the conceptual approach followed which is then implemented in Chapter 3.

2.3 A process for embedding information and knowledge about group activity

This section presents a process to provide a learning group with relevant knowledge extracted from learners' interaction data in CSCL applications for awareness, feedback and monitoring purposes (Caballé et al., 2007f; Caballé et al., 2006; Caballé et al., 2005a). The aim is to greatly improve the effectiveness of the learning exercise.

Here, two difficult problems are to be faced: First, the problem of how to define an efficient process of embedding information and knowledge into a computer-mediated collaboration given that several essential steps need taking into account. Second, how to give relevant and semantically grounded feedback on what is happening in a collaborative learning framework to students and teachers in order to allow them eventually to modify the on-going activity. A solution to the first problem is discussed while providing some suggestions as how to deal with the second.

To manage and provide adequate information and knowledge in a collaborative learning environment, a process is presented consisting of three separate, necessary steps: collection of information, analysis and presentation (see Figure 2.1 and Caballé et al., 2007f). The entire process fails if any one of these steps is omitted. During the first step, a structuring and classification of the generated event information is needed. This information is then analyzed in order to extract the desired knowledge. The final step is to provide users with the essential awareness and feedback from the obtained knowledge. Each of the three stages is next described in turn.

2.3.1 The collection and classification of event information

The most important issue while monitoring group activity is the collection and storage of a large amount of event information generated by the high degree of interaction among the group participants. Such a large amount of informational data may need a long time to be processed. Therefore, collaborative learning systems have to be designed in a way that classifies and pre-structures the resulting information in order, on the one hand, to correctly collect the group activity and, on the other hand, to increase the efficiency during data processing in terms of analysis techniques and interpretations. Due to its importance, the processing step is treated in detail later. As shown in the previous section classification of the information is achieved by distinguishing several high-, mid- and low-level indicators of effective collaboration. Based on this, users' particular actions are further categorized and specified according to the following criteria:

- Who is doing something? (i.e. the originator of the event).
- When did s/he do it? (i.e. timestamp).

- Where does s/he do it? (i.e. the location of the affected resource).
- What is s/he doing? (i.e. the action type and the object involved).
- Why did s/he do it? (i.e. student intentions and motivation which are captured by the indicators associated with each action; e.g., a user performs the action ”create document” in order to generate new information in the shared workspace).

The aim of this stage is to provide a guideline to correctly classify the user actions on the resources during group activity. To this end, a classification process is proposed in which the event information collected from the log files is handled in sequential steps consisting of extraction, identification, coding, and categorization (see Figure 2.2). In particular, first the specific action performed by a user on a resource is extracted (e.g. file document, debate, etc.). Second, this action is then interpreted according to the type of event that was involved in (this information is provided either implicitly by the system according to the context where the action was triggered or explicitly by the user who triggered it). This provides the basic information that is used for the identification of the real intentions or skills shown by the user (e.g. creating a contribution during a debate can be interpreted as either revision or reinforcement of the information depending on whether the contribution was created in the context of a reply or as an observation). Subsequently, the user event is codified taking into account both the user action and the event type by associating a unique code to the user skill identified in the context of the action. Finally, the user event is categorized into one of the group activity indicators defined in Chapter 1 and shown in Figure 2.1.

On the other hand, information from on-line synchronous collaboration is characterized by the spontaneous interactions of their participants. Dealing with this information is a difficult task due to the informality of the participants' contributions, so free dialogue is usually treated by a manual or controlled semi-automated manner. In order to incorporate this kind of information in the automated quantitative process of analysis proposed, the information is to be structured in some way so that it can be collected and processed by computers. To this end, before carrying out a contribution, participants are urged to label their dialogue moves according to the indicators discussed in Section 2. This labeled information is then classified according to the three categories of the model, forming a data source which can be processed and analyzed in a similar way as the asynchronous information (following the processing and analysis steps in Figure 2.2). Note that sometimes participants might not label their contributions correctly which may result in introducing wrong data in the analysis process. For this reason, this step needs human supervision to guarantee the reliability of the information collected.

2.3.2 Efficient processing of the information

Due to the large amount of event information generated in CSCL applications, once this information activity has been correctly collected and classified the issue of demanding computational requirements may come across while processing this information. In order to facilitate this step, CSCL applications may structure this information as log files in a way

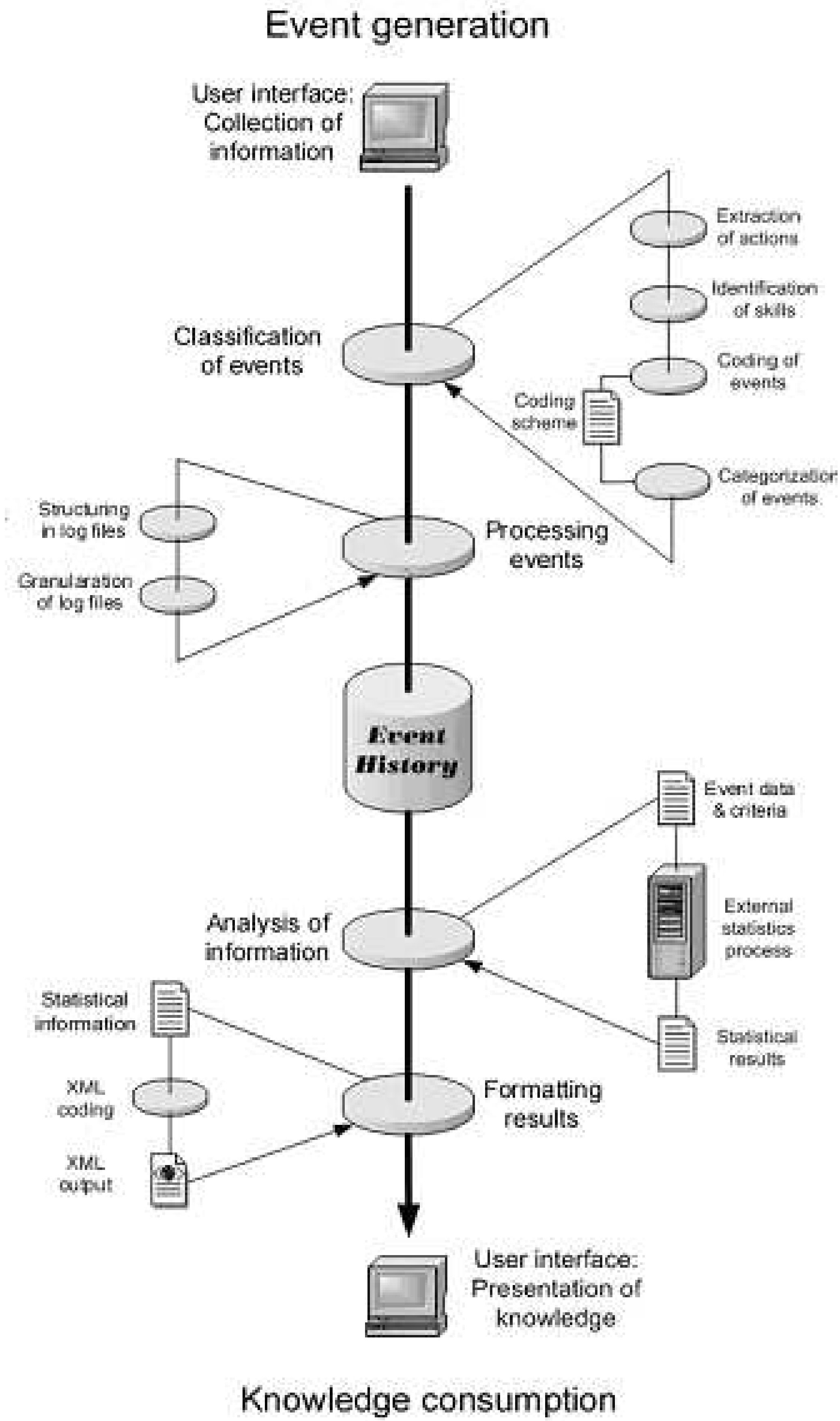

Figure 2.2: The process of embedding information and knowledge into CSCL applications.

that takes advantage of the parallelism in order to process several files (e.g. all the groups in a classroom) at the same time and thus dramatically reduce the overall computational time to process them (Caballé et al., 2008a). As a result, it is possible for these applications to process a large volume of collaboration activity data and make the extracted information available even in real time. Note that this step is optional within this process of embedding information and knowledge and it is pro-posed for efficiency purposes only.

To this end, the following generic steps are proposed so as to correctly structure the event information for later processing (see Figure 2.2, Caballé et al., 2007a, and Chapter 4): by classifying the event information and turning it into persistent data, it is stored in the system as structured files. These files contain all the information previously collected from the system log files in specified fields. These structured files are structured in accordance with certain criteria such as time and workspace, which characterize all group collaboration. The goal is to achieve a high degree of granularity of log files. Thus, during later data processing, it is possible to concatenate several structured files so as to obtain the appropriate degree of granularity (e.g. all groups in a classroom for each 12 hours). This makes it possible to efficiently parallelize data processing depending on the characteristics of the computational resources. To resume, the process of embedding information and knowledge into collaborative learning applications came to the point where the information generated by group activity has been collected, classified and well-structured so that it can be easily and efficiently processed and analyzed during the second stage of this process. This stage is presented next whose purpose is to extract relevant knowledge to be fed back to the participants.

2.3.3 Data analysis and extraction of knowledge

The next stage of this process consists of processing all the information previously collected and classified according to the indicators mentioned before by means of analysis techniques (Caballé et al., 2007f). As a consequence of this analysis, knowledge is generated providing meta-cognition about the state and evolution of interaction, which enhances awareness about the efficiency of group activity, group behavior and the individual attitudes of its members in shared workspaces.

Knowledge extraction is based on criteria related to the three socio-cognitive functions that operate simultaneously during group interaction, namely production function, group well-being and member support and their associated indicators. In that sense, as regards the production function, knowledge is extracted by constantly observing the members' activities (e.g. showing each group member's absolute and relative amount of contributions) or the status of shared resources. In addition, knowledge that is relevant to individual and group well-being can be obtained by exploring the communication and interaction flow among group members (such as members' motivational and emotional state, comparative studies of effective and ineffective groups and so on). Finally, knowledge can be acquired by ill-functioning situations, such as missing or insufficient contributions, lack of participation, etc., which can reveal the need for helping individual members by providing them specific scaffolding where and when this is necessary (i.e. member support).

The definition of a variety of indicators at several levels of description determines the granularity of information to be transmitted to the interested parties. In other words, based on a model of desired interaction (establishing a comparison of the current and desired state of interaction), the analysis approach detects and highlights the indicators which were not satisfied and need to be corrected by redirecting group and individual attitudes. These indicators reveal those aspects of the collaborative learning activity (task performance, group functioning, or scaffolding) that present problems and need to be corrected adequately. Thus, they set up rules and filters in order to extract and summarize only that information which refers to the malfunctioning aspect. The summarized information is finally transformed into useful knowledge that is communicated to and acquired by the group members who use it to improve the performance of the problematic aspect.

Therefore, this approach enables group members to become aware of the progress of their peers in performing the learning exercise both at individual and group level, as well as of the extent to which other members are participating in the collaborative process as this influences their decision making (McGrath, 1991). Moreover, this approach provides tutors with information about students' problem-solving behavior, group processing and performance analysis (Dillenbourg, 1999) for assessment and guiding purposes (Daradoumis et al., 2003). This approach is presented below and constitutes the last stage of the process of embedding information and knowledge into CSCL applications.

2.3.4 Presentation of the knowledge acquired

Here the problem consists in identifying the roles and needs of each learner and the tutor in every moment and being able to decide what information is required to be provided, in which granularity and how to present it. For example, the knowledge obtained from the interaction analysis should be tailored in such a way that the support provided for self-regulation or peer assessment is adapted to the role the learner plays at a particular moment. In that way, scaffolding information would be different for a learner playing a coordinator role from one that plays a software designer role. Moreover, the format used to present the information could vary from case to case. Consequently, several levels are defined that dictate how the acquired knowledge is to be presented, that is, at what format and detail level (Caballé et al., 2007f):

- **Awareness level.** At this level, participants have to be informed about what is going on in their shared workspace, providing information about their own actions or the actions of their peers, or presenting a view of the group interaction, behavior and performance (Gutwin et al., 1995). To this end plane indicator values are displayed that show the state and specific aspect of the collaborative learning interaction and processes that take place. The information presented to the learner can support him/her at a meta-cognitive level.

- **Assessment level.** At this level, data and elements to assess the collaborative activity are provided, so the indicators used are associated with specific weights that measure the significance of each indicator in the assessment process. As in the previous case, the information provided acts at a meta-cognitive level, giving the actors the possibility to evaluate their own actions and behavior as well as the performance of their peers and the group as a whole.

- **Scaffolding (or Guiding) level.** Supporting participants during collaborative activities has become a main concern of current research (Zumbach et al., 2003; Ellis et al., 1991). At this level, information aiming at guiding, orienting and supporting students in their activity are produced. This information is determined by the unsatisfied indicators and helps students to diagnose problematic situations and self-estimate the appropriateness of their participation in a collaborative activity as well as to counsel their peers whenever insufficient collaboration is detected.

In this section the conceptual process is discussed that defines how the embedding of information and knowledge can take place in a CSCL application in an efficient manner and how important it is to consider this knowledge as a means to providing awareness, feedback as well as individual and group monitoring.

Next chapter moves to discuss the implementation of this conceptual approach by means of a computational model based on a generic platform that is used both for the systematical construction of CSCL applications and for embedding information and knowledge from group activity into them. This platform is presented from both the software engineering and knowledge management perspective in the context of the CSCL domain.

Chapter 3

A reused-based CSCL computational model

The development of a clear and well-structured conceptual model such as that seen in the previous chapter constitutes a principled manner for the design of a computational model that implements the process of embedding information and knowledge into a CSCL application. Indeed, the provision of an innovative and effective mechanism that structures and classifies the information into high-level collaborative processes whereas it identifies potential mid- and low-level indicators that measure and evaluate each process. This mechanism contributes and facilitates the building of a portable, general and reusable collaborative learning ontology for the representation, learning and inference of knowledge about each collaborative process. This allows for the design of an effective computational model that reflects and describes task performance, individual and group behavior, interaction dynamics, members' relationships and group support as accurately as possible. In addition, this generic, robust and reusable platform can be used for the systematic construction of CSCL applications endowed with enriched capabilities for providing more efficient knowledge management and scaffolding (awareness, feedback and group monitoring). This chapter is going to focus on the former issue while the following chapters address the latter. To this end, in this chapter, an in-depth research is provided first on the following issues Generic Programming, Service-Oriented Architecture, and Model-Driven Architecture. This study becomes the very rationale of the CSCL platform presented in this chapter.

3.1 Generic Programming

In traditional forms of engineering, productivity, quality and cost are such important factors in industrial processes that the very survival of companies depends upon them. For this reason, great efforts have been made to improve the techniques, methods and tools which are available for product development and the results have clearly been spectacular. However, in the case of software development similar progress has not been made so far even though there is no essential difference with other forms of engineering. Indeed, the fact that the technology is more recent and the product is highly complex may explain the reasons, but the key is doubtless to be found in the fact that the concept of reusability has not been sufficiently exploited.

The reuse of previously created product parts leads to reduced costs and improved productivity and quality to such an extent that industrial processes will take a great leap forward. In all advanced forms of engineering new products are usually developed by reusing tried and tested parts but in software engineering it seems that new products are almost always developed from scratch. To benefit from the advantages of reusability it is necessary to develop better methodologies such as Generic Programming (GP) that facilitate this possibility. GP when applied in the context of computer software development is an innovative paradigm that attempts to make software as general as possible without losing efficiency. It achieves its goal by identifying interrelated high-level family from a common requirement set. By the application of this technique, especially in design phases, software is developed offering a high degree of abstraction which is applicable to a wide range of situations and domains.

3.1.1 Objectives

By applying GP to develop computer software important objectives are achieved (Caballé and Xhafa, 2003b; Caballé, 2008d):

- **Reuse**. This is the main reason for applying GP ideas. The real objective, however, is to be able to reuse and extend software components widely so that it adapts to a great number of interrelated problems. This concept of reutilization is much broader than has been seen until now.

- **Quality**. Due to the great potential for reutilization of GP, it is necessary to guarantee maximum quality. Here *quality* refers to the correctness and robustness of implementation which provides the required degree of reliability. Furthermore, GP provides innate reliability as the implementations are nothing more than skeletons, without details, and as such are simpler to construct during the codification phase and hence can be produced with minimal errors.

- **Efficiency**. As in the case of the quality, the efficiency of components is to be guaranteed as if this not done the performance repercussions will be noted, just as with lack of quality, in all of the systems involved.

- **Productivity**. Inherent to reutilization is the saving through not having to create software components again that already exist. Hence, there is an increase in computing production, which is one of the benefits that GP seeks in order to bring the production potential of computing applications as in other industries.

- **Automatisation**. Here the aim is to automatise the processes so that general requirements with a high level of abstraction and specially designed tools can be used to produce operative programmes. GP provides the skeleton for the initial generic requirements. The resulting programme will be the base for other more specific programmes which in turn will construct others following a cascade process thus giving a more solid base permitting more automatization.

- **Personalization**. As the general requirements are made more particular, so the product that is generated becomes more optimized to meet the specific needs of the client. GP intervenes largely in a first phase of this abstraction/specialization/ personalization cycle by identifying the abstractions and supplying a first level of specialization; this is a great challenge of the software industry.

GP also represents one important technique to achieve effective Product Lines following the Product-Line Architecture(PLA) approach (Czarnecki and Eisenecker, 2000). PLA promotes developing large families of related software applications quickly and cheaply from reusable components. In PLA, a certain level of automation is provided in the form of generators (also known as component configuration tools) to realize solutions for large parts of the systems being developed.

3.1.2 An application example

The development of applications based on GP not only has the practical advantage of improved productivity but also results in software that is more robust and reliable. Considerable research has already been done into the construction of libraries of generic data structures based entirely on GP (Standard Template Library (Musser et al., 2002) and Data Structures and Algorithms in Java (Goodrich and Tamassia, 2001) but very little has been done into the development of computer systems.

In order to validate the approach, previous research (Caballé and Xhafa, 2003a; Caballé and Xhafa, 2003b) considered the construction of a software platform called General Purpose Library (GPL) [1]. See Appendix B for a technical overview of this platform. The GPL is made up of components of greatly generic use that creates the skeleton for the construction of complex systems requiring the management of the users interacting with the system and optimization of the system's resources. The aim of this study was to investigate the feasibility of the design of generic software as the basis of complex and extensive domain computer systems. The library was based on the GP paradigm as the aim was to encourage the greatest possible reusability of its own generic components for the development of the specific computer systems highly complex. The first step was to identify those parts which are common to most applications of this domain and then proceed to isolate the fundamental parts in the form of abstractions from which the basic requirements were obtained. Once a logical division into components and subsystems was made, each component was analyzed and designed separately employing OO methodology. In order to maintain intact the ideas of GP design that were found, an implicit logical layer was implemented that creates a correspondence between the GP design and the OO design. Since Java has a great predisposition to adapting and correctly transmitting a high degree of abstraction and make the software reusable (Caballé and Xhafa, 2003a), this library was implemented in this programming language. As a result, the advantages that GP offers with regard to quality, efficiency, productivity and so on provides a solid basis for the construction of specific software that is faster, more efficient and highly reliable.

[1] see API at: http://cv.uoc.edu/~scaballe/tfc/api (Web page as of April 2008).

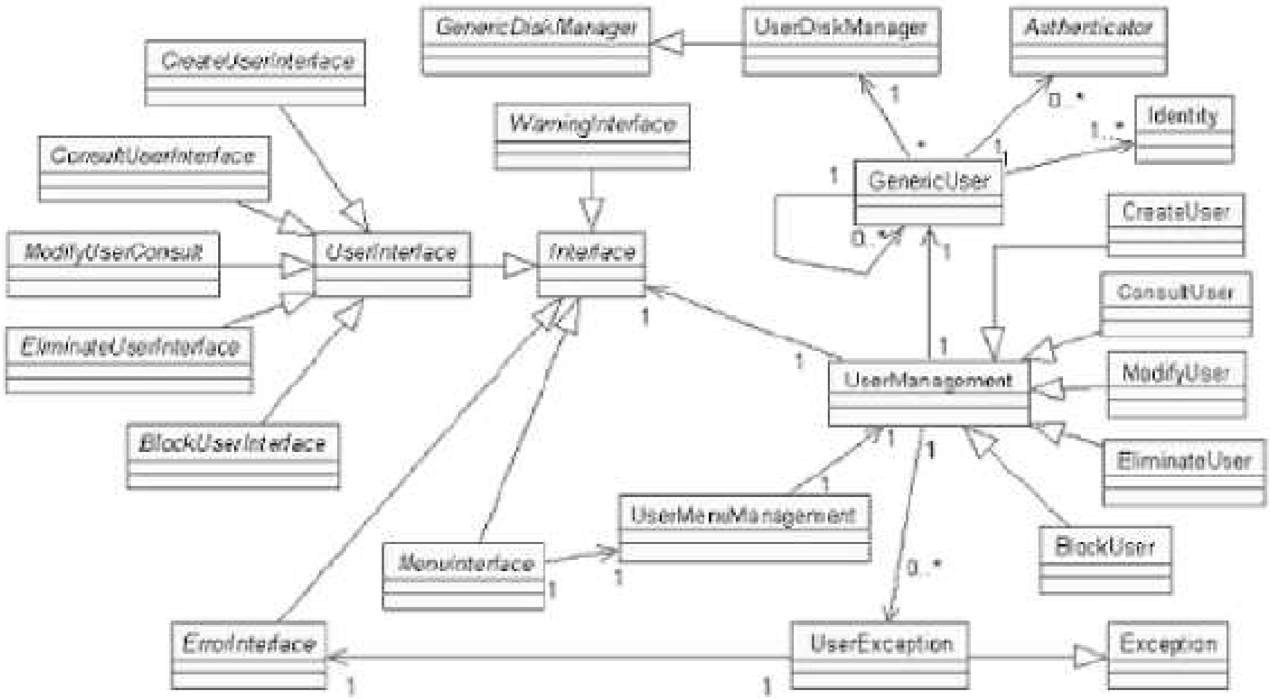

Figure 3.1: Class diagram for GenericUserManagement subsystem

The GPL library (Caballé and Xhafa, 2003b, Caballé and Xhafa, 2003c) is made up of three components which constitute the skeleton of the basic structure of whatever application is constructed using this library.

- The **Users** component. It is made up of two subsystems:
  - The *GenericUserManagement* subsystem administer the basic elements participating in the administration of the users of a computing system. The concept of generic user represents a person, group, device, system, organization etc., which is able to have multiple identities. So, for example, if the user is a person, the identities can be the name of the person, his/her national identity number, his/her social security number etc. Each identity represents a role of the user in a specific setting, so, the same user is able to have multiple roles. In the same way an identity is unique both at user level and within the group of all the users of the system. A generic user may also have multiple authenticators which are both public and private so as to validate each one of the aforementioned identities. The private keys of the authenticators will be encrypted for security, leaving users free to choose and implement the encryption algorithm that best fits their particular needs. So, the Generic User entity will collect all the information that is available about a user of a computing system so as to be able to use it in the construction of most computing applications involving user interaction (see Figure 3.1).
  - The *UserProfileManagement* subsystem is designed to manage user profiles with information regarding the users' personalization of the setting (or *environment*). By *user profile* it is understood the information which specifies how a specific user performs within the system. This information is obtained through the elements of

the profile where each element specifies a particular piece of information about the user (alias, language, font and letter size, colors, etc.), which is modeled as an ElementProfile. The sum of all of these ElementProfile forms the UserProfile. Each user may belong to one or more user groups in order to share information (e.g. a team carrying out a project, a set of devices sharing the same data, etc.).

- The **Security** component. It is made up of two subsystems containing a full description of the measures taken and rules adopted in order to protect the system resources against the deliberate or accidental ill use by the users and to protect against unauthorised access to the system.

  - The *AuthenticationManagement* subsystem is designed to identify the user wishing to enter the system and manages the start up and close down procedures of users recognized by the system. The generic process of authentication consists of starting a *UserSession* in which, firstly, an uninitiated user session is created where the user's authentication credentials are required. Users present their Identity together with their Authenticators as entry parameters. From the identity, the *GerericUser* is identified and the Authenticator stored in the system is compared with that which has been supplied by the user and its validity is checked.
  - The *AuthorisationManagement* subsystem has the main aim of administering the system's security code. This code is made up of all the norms and rules of security established in the system and amounts to the system's security policy. *Security policy* means all of the norms and declarations specified to determine entry to valuable resources of the system. This information will essentially arise from permission granted to users to enter specific resources through specific actions. The aim is to limit entry to the system's valuable resources to users. Granting permission to a specific user will entail the assignation of a privilege level to the user in question based on the existing information. These privileges will be assigned during the start up of the session once the user has successfully been authenticated.

- The **Control** component is made up of two subsystems containing those devices (e.g., log file) and processes (e.g., calculation of statistics) used in the control and maintenance of the system to administer the available resources correctly. It aims to improve both performance and security.

  - The *MonitoringManagementSystem* has the main aim of administering all of the data produced by the system itself as a result of the events occurring during normal use. The key entity of this subsystem is the *History* made up by the file log which records all movements and incidents in the day-to-day operations of the system such as the start up and close down of user sessions, entry of resources, failed authentication, and system errors. Each one of these events is recorded in an *ElementLog* entity containing the date and time the event occurs together with a textual description. So, the *History* entity contains all information affecting the system generated by its normal use (see Figure 3.2).
  - The *PerformanceControlManagement* subsystem has the main aim of administering and maintaining the statistical studies resulting from the system's own data from events generated by the functioning of the system

and stored in the *History*. The *Statistics* entity will provide the ability to extract useful information about these events that will lead to determine the level of performance and reliability of the critical parts of the system. One *Statistics* entity will contain the data of the History which, together with a Criteria entity, will act as a source of reference for the extraction of the information required. The *Criteria* entity may be multiple taking in different attributes or fields of the *History* so as to construct complex consultations that will give a detailed knowledge of the workings of the system. The statistical results may be offered in different formats depending on the system's needs and ability to visualize including text form, html, graph (bar charts, pie charts, etc.) and other available formats. These results will make it possible to elaborate detailed reports into the state of the system that will show the way to making appropriate corrections to system deviations and to carrying out any necessary improvements.

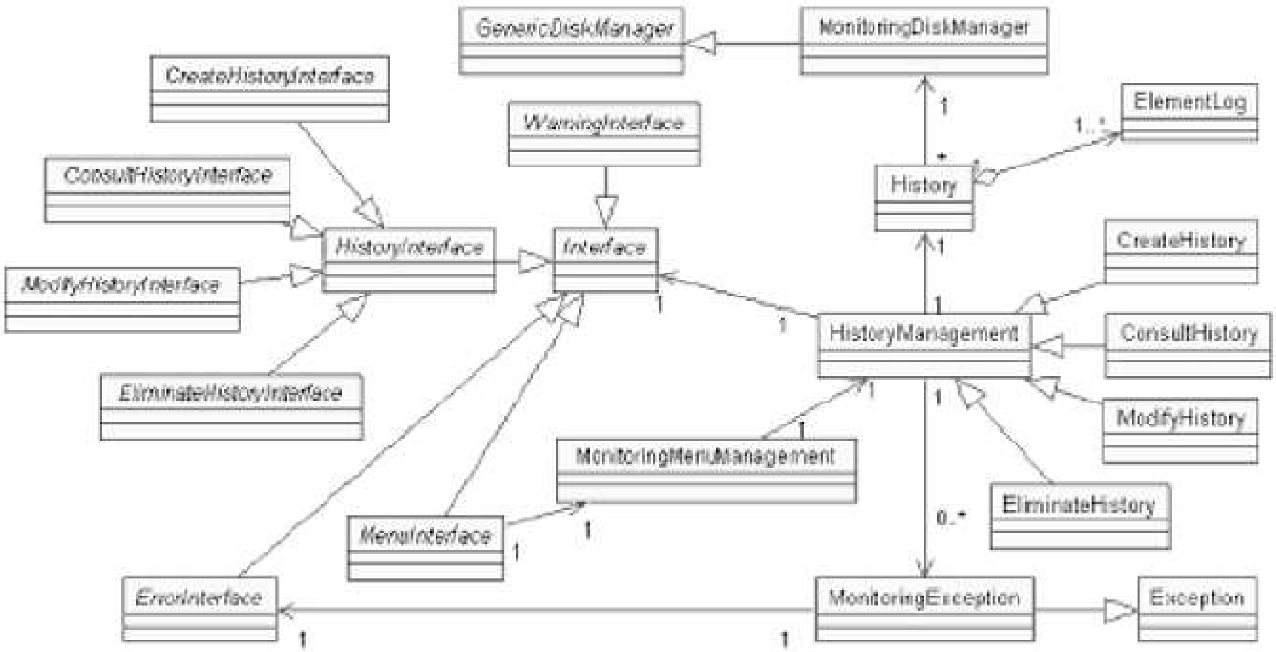

Figure 3.2: Class diagram for MonitoringManagementSystem subsystem

The *user interface* is generically focused so as to make particularization in graphic and text modes possible. Even though, in these environments, the user interface will typically be in graphic mode, it is necessary to consider this abstraction in order to make the logical part of the application independent from the specific design of the graphic user interface. The design of the persistence is also generic and so a *disk manager* abstraction has been considered to act as a bridge between the future application and its data to make the design of the persistence independent from the specific technology that will manage the data and allow the treatment of both ordinary text files and the database during particularization. Finally, a complete hierarchy of exception provides a high degree of reliability without depending on the error treatment of the specific platform supporting the software.

This approach may be applied to a very large number of potential applications making it possible to be benefit from reuse a great deal. Indeed, the following requirements are found extremely often while facing a new application and may be met by using the GPL library:

- Management of users, their profile and workspaces as well as the system resources.
- Login/logout mechanisms to validate users. Protect the system from both the unknown users and the intentional or accidental bad use of its resources.
- Identification and the notification of recent events and the consultation of the system events history in order to extract statistics about any time period of the system's life.

During an application construction that matches these requirements, it can be checked how even in design phases the GPL library adapts perfectly thanks to a good matching of those generic entities proposed with the instantiation of those made here. For example, how *GenericUser* matches with whatever user type is found in a any application (single users, groups, etc.) and *Identity* becomes the unique identifier of the user (e.g., e-mail address, social security number, etc.) and also how *ElementLog* gathers all the information needed to identify the system's events from which it is possible to carry out their notification to the rest of the users. With regards to functionality, the generic processes are also well-matched with specific ones such as *AutenticationValidation*, which once instantiated, permits carrying out user authentication and accessing to their own workspace and resources. The *user interface* is instantiated in a graphic mode taking advantage of the genericity of the library. Regarding the persistence, *GenericDiskManager* abstraction is available and its specialisations which represent a bridge that keeps the logic of the application independent from its data allowing for persistence in different models.

Next sections take this generic approach so as to present a particularization of the GPL platform as a fully reused its components in the CSCL domain. The outcome of the approach is a generic, highly reusable platform for the systematic construction of collaborative learning applications.

3.2 A generic gaze at the collaborative learning applications

In this section, a generic view of the CSCL domain is given by analyzing and taking into account the commonality found in the requirements of most of collaborative learning environments and, in particular, the provision of efficient knowledge to CSCL applications.

3.2.1 Common basis for CSCL applications

In the last years there has been an explosion of new CSCL applications aiming to create collaborative learning environments where students, teachers, tutors, etc., are able to cooperate with each other in order to accomplish a common learning goal. To achieve this goal, the collaborative applications must provide support to three essential aspects: coordination, collaboration and communication; with communication being the base for reaching coordination and collaboration

(Caballé et al., 2004). Collaboration and communication might be *synchronous* or *asynchronous*. The former means co-operation at the same time with typically fine-grained notifications giving immediate feedback about the activities of other participants whereby the shared resource (such as a text document and a message) will not have a lifespan beyond the sharing. The latter means cooperation at different times and the shared resource will be stored in a persistent support.

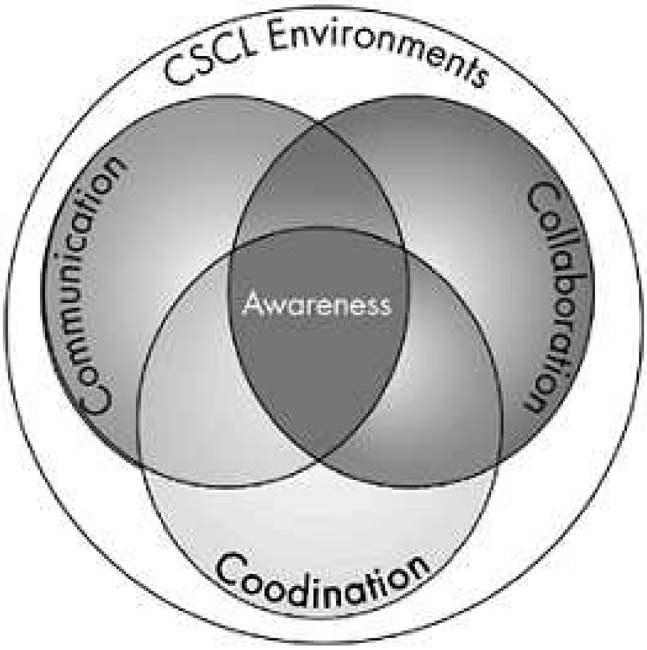

Figure 3.3: The essential aspects in any collaborative learning application.

The different areas overlap each other (see Figure 3.3) and any collaborative system must support all the three aspects (Ochoa et al., 2002; Caballé et al., 2004):

- *Coordination* is an important aspect of any collaborative activity. It entails the combination and sequencing of otherwise independent work toward the accomplishment of a larger goal. In a collaborative learning environment, coordination mostly refers to the tasks toward the learning group formation and the definition and planning of the group objectives. Moreover, the group coordinator may track task status, deadlines, resource usage, working results, or other critical process parameters to correctly lead the group.

- *Collaboration* relies on students sharing all kind of documents. The sharing of resources between several participants is therefore a central functionality of CSCL systems. Sharing may be synchronous, with several participants accessing the same resource at the same time (that is, they work on the same copy of the document), or asynchronous, with different participants accessing the same resource at different times (each of them works on a different copy of the same document).

- *Communication* is another functional aspect of collaboration systems aiming to support the communication between two or more collaborative learning participants. Communication includes text messages, spoken interactions, or non-verbal exchanges like gestures in a video conference (Baloian, 2002). Communication may take place asynchronously (different participants communicate at different times such as email, debate, etc.) or synchronously (participants communicate at the same time such as chat, video conference, etc.). The communication support is based on four elements involved: a message as the information carrier between a sender process and a recipient process (which receives and possibly process the message) through a channel (Ochoa et al., 2002). Moreover, in this context, it is necessary to implement different ways of message addressing such as point-to-point, multicast and broadcast.

- *Awareness* is essential for any of the three forms of cooperation seen above. It allows for implicit coordination of collaborative learning, opportunities for informal, spontaneous communication and gives users the necessary *feedback* about what is happening in the system. In particular, on the one hand, synchronous awareness lets users know exactly what other co-participants are doing (e.g. during a shared editing session shows who is editing what) and when documents are in use by others. On the other hand, asynchronous awareness determines who, when,, how and where shared resources have been created, changed or read by others.

In order to improve the collaboration within a group it is important to take into account both current and future behavior of all user types and the fact that user objectives or intentions may change as they interact with the system. To that end, it is essential to design some kind of user and group models describing, for example, the user characteristics, intentions, beliefs, knowledge, skills, roles and collaborative activities (Caballé et al., 2004; Caballé et al., 2007d). Moreover, the user and group models should be open enough to let add new services and collaborative activities to them according to the participant needs.

The design of the CSCL user interface offers many more challenges than the design of interfaces for single user applications (e.g. multi-user editors). The user interface must provide information about what others are doing to efficiently support collaborative tasks and additional information has to be presented. The latter refers to the effects of other users' activities which must be communicated by visual or audio signals. Therefore, the user interface is the main way to support awareness in multi-user collaborative environments.

Although most research efforts in CSCL areas have been dedicated to developing distance learning environments, most learning activities still take place in the traditional face-to-face classroom (Balonian et al., 2002). To that end, a generic approach such as that of this thesis should support the common basis from both scenarios and it is possible to instantiate CSCL applications both for virtual learning (i.e. most of participants are physically in different places) and for traditional learning (i.e. all the participants are physically found in the same place, usually in a classroom). In this thesis, though it mostly refer to virtual CSCL environments, the principles are the same for both scenarios.

3.2.2 The importance of interaction analysis in CSCL practices

Collaborative learning environments are characterized by a high degree of user-system interaction thus generating a huge amount of action events. The management of action events is a key issue in applications since, on the one hand, the analysis of data gathered from real life on-line collaborative learning situations would help important issues in group functioning and collaborative learning process to be understood; In addition, this can guide both the design of more functional workspaces and software components and the development of better facilities such as awareness, feedback, monitoring of the workspace, assessment and tracking of the group's work by a coordinator, tutor, etc. Indeed, by filtering out the data, an adequate event management allows for the establishment of a list of parameters that can be used for analyzing group space activities (e.g., tutor-to-group or member-to-member communication flow, asynchronism within the group space, etc.). These parameters would allow the efficiency of group activities to be improved and group behavior and individual attitudes of its members in the shared workspace to be predicted.

Furthermore, in designing applications it is necessary to correctly organize and administer both the resources offered by the system and the users accessing these resources. All of this user-resource and user-user interaction generates events or logs which are collected in log files and represent the information basis for the performance of statistical processes aimed at obtaining useful knowledge of the system. This will facilitate the collaborative learning process by keeping users aware of what is going on in the system (e.g. the contributions of others, the new documents created, etc.) and controlling users' behavior in order to provide them with support (e.g. helping students who are not able to accomplish a task on their own). Therefore, user-user and user-resource interaction is crucial in any learning collaborative environment to make it possible for groups of students to communicate with each other and to accomplish common objectives (e.g. a collaborative classroom activity).

Although user interaction is the most important point to be managed in applications, it is normally also important to be able to monitor and control the performance and general functioning of the system. This will enable the administrator to continuously track the critical parts of the system and act if necessary. Furthermore, this adds an implicit security layer to that which already exists (e.g. controlling users' habits making it possible to detect fraudulent use of the system by unauthorized users).

In order to efficiently communicate the knowledge achieved from group activity to users in terms of awareness and feedback, CSCL applications must provide full support to the above-mentioned three essential aspects existing in any collaborative application, namely coordination, communication and collaboration (see Section 3.2.1 and Figure 3.3) so as to create virtual environments where students, teachers, tutors, etc., are able to cooperate with each other in order to accomplish a common learning goal. *Coordination* involves the organization of the group in order to accomplish the objectives set and the monitoring of user activity, which is possible by maintaining the awareness of participants. *Communication* is related to the communication support basically by messages among the users within and between groups and can be in both synchronous and asynchronous modes. Finally, *collaboration* lets group members share any kind of resources, which is also in both synchronous and asynchronous modes. Both coordination and collaboration as well as communication will generate many events which will be communicated to the users after these events have been handled and analyzed in order to provide users with as much immediate awareness as possible and a constant flow of as much feedback as possible.

3.3 A computational model for data analysis and management in CSCL

The main contribution of this thesis is a generic, reusable, robust, flexible, interoperable, component-based and service-oriented platform called Collaborative Learning Purpose Library (CLPL)[2] (Caballé et al., 2004; Caballé et al., 2006; Caballé et al., 2007e; Caballé et al., 2007f; Caballé et al., 2007g; Caballé et al., 2008d). See Appendix B for a technical overview of this platform. The CLPL is based on the Generic Programming paradigm so as to enable a complete and effective reutilization of its generic components as the skeleton for the construction of any collaborative learning application. This generic platform implements the conceptualization of the fundamental needs existing in any collaborative learning experience. In the context of this research, this platform is especially used as a computational model to embed information and knowledge from group activity CSCL applications as the implementation of the conceptual model for data analysis and management described in detail in Chapter 2.

For this reason, special attention is paid to event analysis and management in developing the CLPL. To this end, a specific component is conceived so as to make it possible to collect, structure and classify all the generated event information as well as to process this information for its later analysis. This component completely specifies and implements the first two stages of the above-mentioned information and knowledge embedding process. The third stage, presentation (for user awareness and feedback), is fully accomplished by a different component, which implements the three basic elements involved in any groupware application (i.e. coordination, communication and collaboration) and is especially responsible for communicating the obtained knowledge to users in terms of awareness and feedback.

In order to meet these requirements, the development of the CLPL is based on the Model-Driven Development (MDD) paradigm and the framework supporting it, namely Model-Driven Architecture (MDA) (Czarnecki, 2005). In proposing MDA, the CLPL development takes advantage of two key ideas that have had significant influence in addressing the current challenges in software development (Caballé, 2008d): Service-Oriented Architectures (SOA) and Product Line Architectures (PLA). As to the former, SOA provides great flexibility to system architectures by organizing the system as a collection of encapsulated services. Hence, SOA relies on services which represent the behavior provided by a component to be met and used by any other components based only on the interface contract. As to the latter, PLA promotes developing large families of related software applications quickly and cheaply from reusable components. In PLA, a certain level of automation is provided in the form of generators (also known as component configuration tools) to realize solutions for large parts of the systems being developed (Czarnecki and Eisenecker, 2000). Taking these approaches into consideration, the CLPL is based on SOA and the Generic Programming paradigm (Czarnecki, and Eisenecker, 2000; Caballé, and Xhafa, 2003b) as the central part of the development in MDD.

In particular, in developing the CLPL, a Platform Independent Model (PIM) was first created by applying the following Generic Programming ideas (see Chapter 1 and Caballé and Xhafa, 2003b): (i) define the semantics of the properties and domain concepts, (ii) extract and specify the common and variable properties and their dependencies in the form of abstractions found in the CSCL domain, and (iii) isolate the fundamental parts in the form of abstractions from which

[2]Last release of the CLPL is version 1.1, which can be found at: http://clpl.uoc.edu/download/CLPL1.1released.zip (Web page as of April 2008)

the basic requirements were obtained, analyzed and designed as a traditional three-layer architecture (i.e. presentation, business and information). To this end, first, the PIM was expressed using UML as the standard modeling language promoted by the OMG (see Figure 3.4). Second, two different Platform Specific Model (PSM) have been constructed so far from the PIM: A Java implementation in the form of a generic component-based library and a service-oriented approach by using Web-services technology.

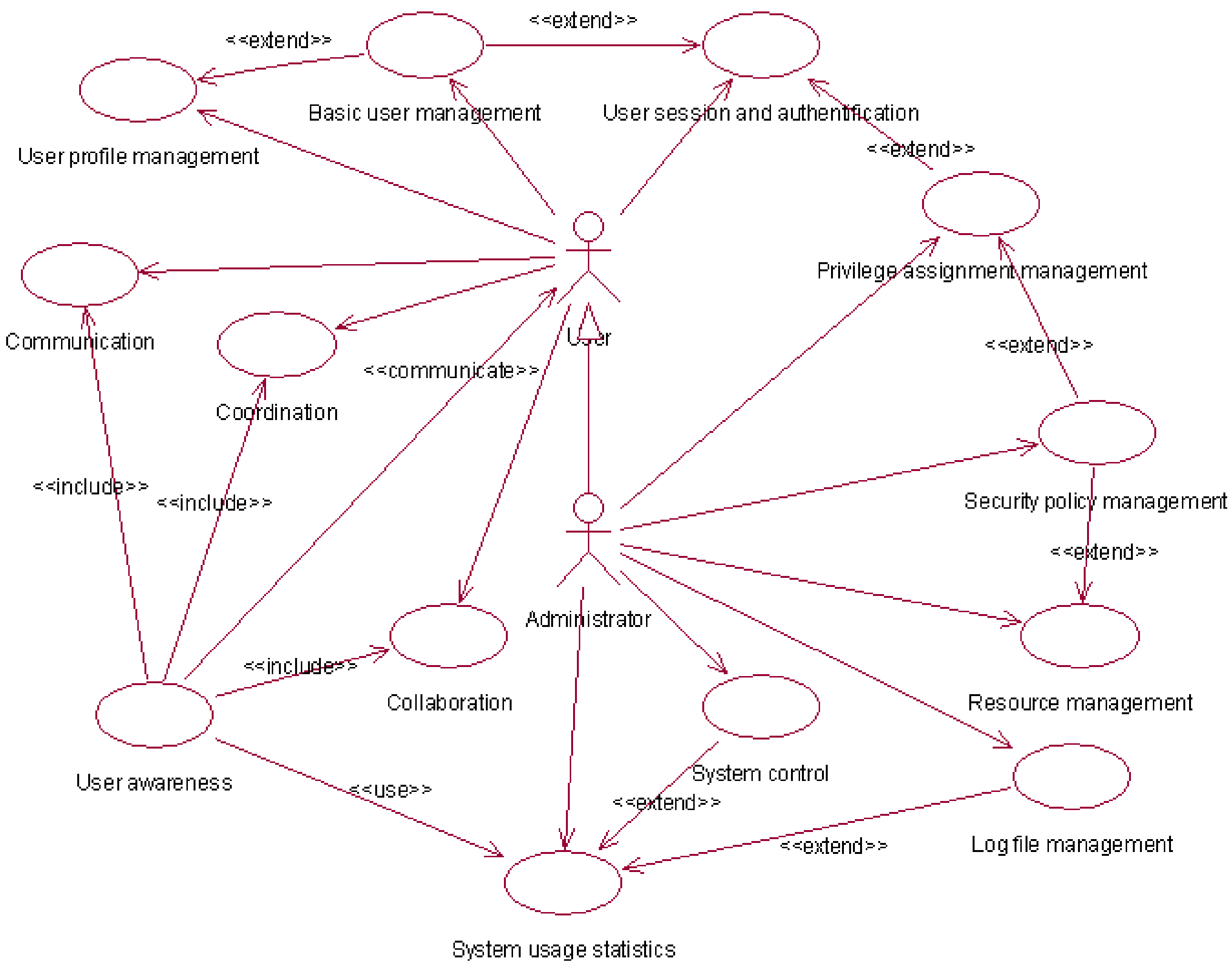

Figure 3.4: UML-based use case diagram with the general requirements of the CLPL.

The ultimate aim of the CLPL is to enable a complete and effective reutilization of its generic services and components as the skeleton for the construction of any collaborative learning application, and in particular CSCL applications. Thus, this platform implements the conceptualization of the fundamental needs existing in any collaborative learning experience. In addition, the CLPL is highly interoperable in distributed environments permitting complete flexibility of the services offered in terms of implementation languages and underlying software and hardware platforms.

For the rest of this section an UML-based PIM model for the CLPL is described in detail by means of, first, a general view of the CLPL architecture and then an in-depth study is performed of the part of the CLPL architecture that supports the embedding of information and knowledge about group activity into CSCL applications (the complete PIM of the

CLPL is found at http://clpl.uoc.edu/docs/CLPLdevelopment.pdf). Next section faces the PSM approach by incorporating specific technology to the CLPL.

3.3.1 The CLPL architecture

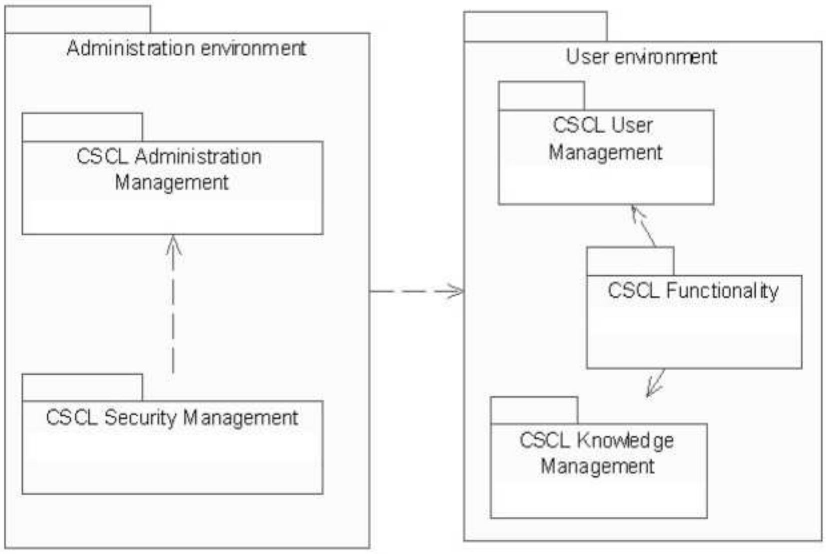

Figure 3.5: Graphical representation of the CLPL components.

The CLPL (Caballé et al., 2007e) is made up of five components (see Figure 3.5) handling user management, administration, security, knowledge management, and functionality, which map the essential issues involved in any collaborative learning application.

- CSCL User Management component: this contains all the behavior related to user management in applications, which can act as a group coordinator, group member, group-entity and system administrator. It will tackle both the basic user management functions in a learning environment (namely registration, deregistration, modifications, joining a group, or meeting group members) and the user profile management. The latter implements the user and group models within a collaborative environment, thus this component provides the generic ProfileElement entity which dynamically allows new user and group needs to be met.

- CSCL Security Management component: this contains all the generic descriptions of the measures and rules decided upon to resolve authentication and authorization issues and so protect the system from both unknown users and the intentional or accidental ill use of its resources. Its genericity lets programmers implement these issues with the latest cryptographic security mechanisms.

- CSCL Administration Management component: this contains the specific data from log files and those analyzes (i.e. statistical computations) required to perform all the system control and maintenance for the correct administration of the system and to improve it in terms of performance and security. Moreover, it will manage the resources of the collaborative workspace, which can be managed by a group member acting as an administrator within the group.

- CSCL Knowledge Management component: this manages all the specific and large user events in order to handle the data of user interaction as crucial information for the extraction of the essential knowledge to notify users of what is going on in the system as well as to monitor user behavior and control system resources. To this end, this component has been split into the *CSCL Activity Management* and *CSCL Knowledge Processing subsystems*. The former aims to collect and classify the user events captured according to a complete hierarchy of user events (see Fig. 4) provided, which is based on the above-mentioned three generic group activity parameters: task performance, group functioning (i.e., interaction behavior) and scaffolding. The latter is responsible for the performance of the statistical analysis of the event information previously handled and includes another generic hierarchy (see Fig. 7) that contains those statistical criteria which are most common in these environments (e.g., the number of students connected over a period of time, the average student working session). Furthermore, it will enable log information to be exported and extracted in different formats for later statistical analysis in external statistical packages. The final objective of this component is to extract valuable information from the events generated with the aim of revealing useful knowledge. Since this component represents the core of thie CLPL, it is explained in great detail later on in the next sections.

- CSCL Functionality component: this forms, along with the previous component, the basis of the collaborative learning environments by defining the three basic elements involved in any groupware application (see Figure 3.3), namely, coordination, communication and collaboration (Caballé et al., 2004, Caballé et al., 2007e). The different areas overlap each other, and any collaborative system must support all three aspects (see Figure 3.3). Due to their importance, this component provides several subsystems or modules so as to provide direct support to each of these areas, namely *CSCL Coordination*, *CSCL Communication* and *CSCL Collaboration*. The coordination support module offers the basic tools to facilitate group organization in planning and accomplishing the members' objectives as well as group monitoring by modeling the awareness of its participants. The communication support module involves four basic elements, the sender, message, channel and receiver (Ochoa, 2002), and can be implemented in several ways depending on the means of message transmission (point-to-point, multicast and broadcast). Moreover, each message can be delivered asynchronously (as in the case of an email, where the message is made persistent by default) or synchronously (as in a chat, where conversation is made persistent so that it can later be processed). Finally, the collaboration support module lets members share both software and hardware resources in both synchronous (e.g. real-time editors) and asynchronous (e.g. file sharing) modes. This component also supports the presentation of the information managed by the CSCL Knowledge Management component by means of a subsystem called *CSCL Awareness* with the aim of providing participants with immediate awareness of what is going on in the group. Furthermore, in the last few years, feedback (Zumbach, at al, 2003) is receiving a lot of attention due to its positive impact in on-line collaborative learning in such areas as group motivation, interaction, or problem-solving abilities (Caballé, at el, 2005b).This characteristic is also supported in this component by another subsystem called *CSCL Feedback*, which also takes advantage of the knowledge extracted from the group activity to provide participants with a constant flow of as much feedback as possible. This component is also described further in the next sub sections.

These CLPL components can be directly reused in the construction of specific efficient, robust, multiplatform and reusable CSCL environments, especially Web-based applications (Bentley et al., 1997; El Saddik et al., 2001). As mentioned above, these components in turn fully reuse a more generic library, the above described Generic Purpose Library (GPL) (Caballé et al., 2003b), whose domain requirements are a high degree of user-user and user-system interaction as well as optimal management of system resources. By including collaborative learning domain into the GPL domain, the CLPL will represent a particularization of the GPL (see Figures 3.6, 3.7 and 3.8 as examples of reusing the GPL for the development of the CLPL)

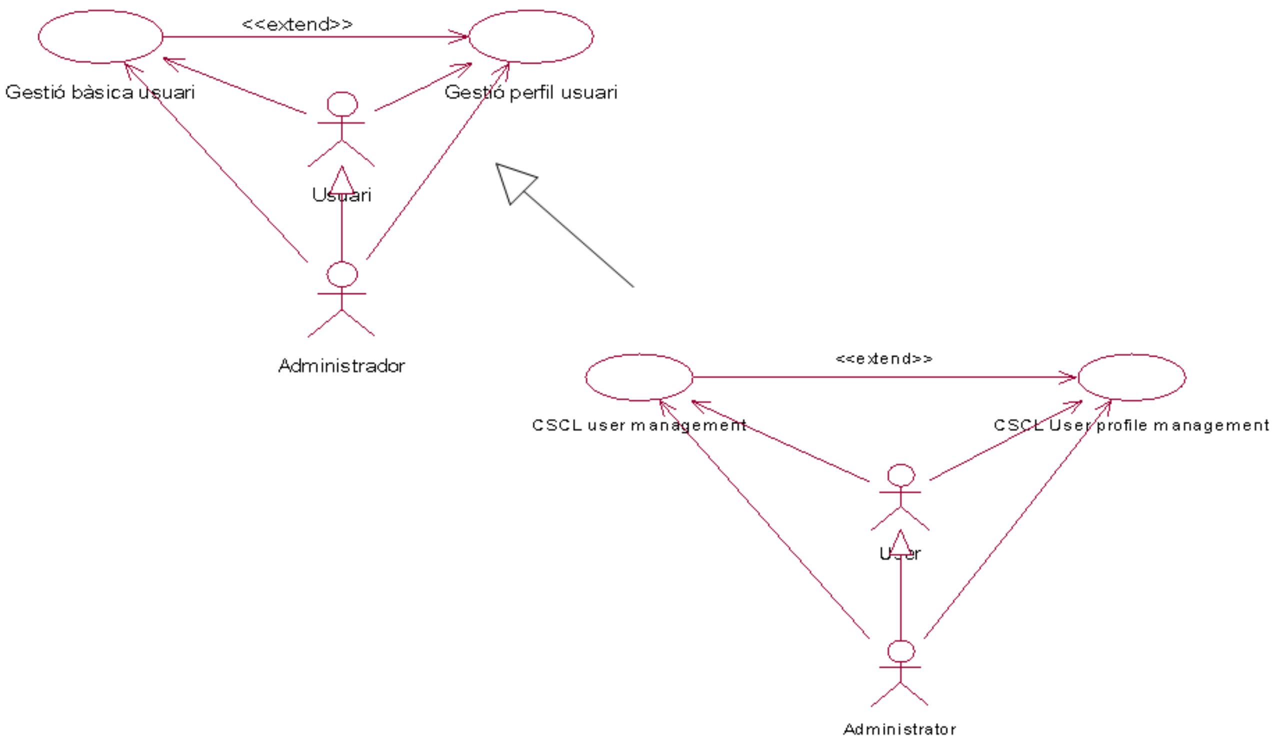

Figure 3.6: Reusing at requirements level as part of the PIM model of the GPL.

The GPL provides the CLPL with the following and important behavior by default:

- In order to improve collaboration within a group, it is important to take into account both current and future behavior of all user types and the possibly changing objectives and intentions of the users as they interact with the system. To this end, generic user and group models have been designed to describe the users' characteristics, intentions, beliefs, knowledge, skills, roles and collaborative activities amongst others. Moreover, the user and group models are sufficiently open as to allow new services and collaborative activities to be added in accordance with the needs of the participants.

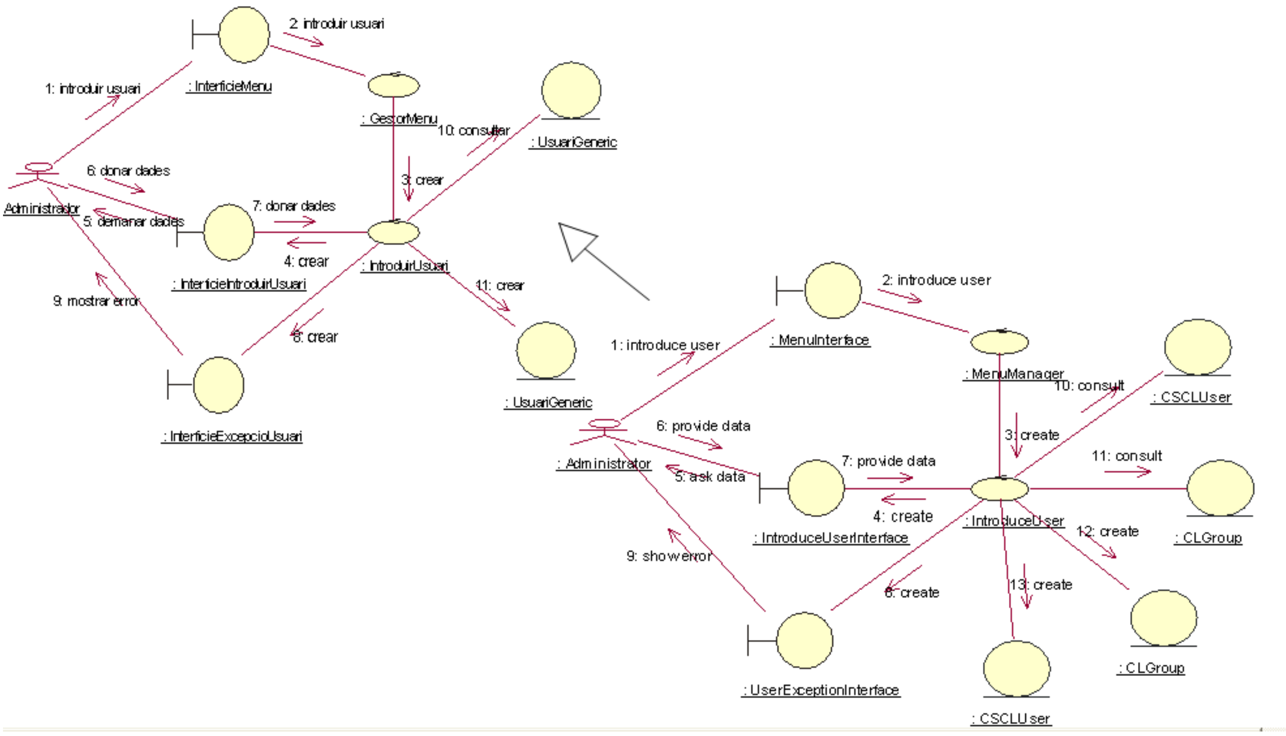

Figure 3.7: Reusing at analysis level as part of the PIM model of the GPL.

```
package clpl.users;

import gpl.users.*;
/**
 *  Abstract class carrying out the logic part of the group creation within a CSCL environment.<br>
 *  @author <a href="mailto:scaballe@uoc.edu">Santi Caballé Llobet</a><br>
 *  @see <a href="http://cv.uoc.edu/~scaballe/gpl/api/gpl/users/GenericUserCreation/">gpl.users.GenericUserCreation</a><br>
 *  @version 1.0a; 15-12-2003; java version "1.2.2"<br>
 */
public abstract class CLGroupCreation extends GenericUserCreation{

    /**
     *  Constructor without parameters from the class CLGroupCreation.<br>
     *  Creates a group.<br>
     *  (pre: the group does not exist in the data store.)<br>
     *  (post: the group has been created in the data store.)<br>
     */
    public CLGroupCreation() throws GenericUserException, Exception {
    }
}
```

Figure 3.8: Reusing the Java-based PSM of the GPL.

- The design of the user interface in collaborative learning applications offers many more challenges than the design of interfaces for single user applications (e.g. multi-user editors). The user interface must provide information about what others are doing to efficiently support collaborative tasks, and awareness information regarding the effects of other users' activities has to be communicated by visual or audio signals. The user interface is therefore the main way to support awareness in multi-user collaborative environments. Furthermore, the user interface is generically focused so as to make particularization in graphical and text modes possible. Even though the user interface in Web-based collaborative learning environments will usually be in graphic mode, it is necessary to consider generic focusing in order to make the logic part of the application independent from the specific design of the graphic user interface.

- The design of the persistence in the CLPL is also generic and thus a disk manager abstraction has been considered. The disk manager acts as a bridge between the future application and its data to make the design of the persistence independent from the specific technology that will manage the data. This way, it is possible to treat both ordinary text files and different database system managers during particularization. Furthermore, a complete technology-independent conceptual data model is provided as part of the PIM (see sub section 3.3.3 and Figure 3.12), which may be realized in different technologies managing generic persistence.

- Robustness is offered through a complete hierarchy of error treatment and so a high degree of the component quality and reliability is guaranteed without depending on the error treatment of the specific platform supporting the software.

The whole CSCL activity is performed in the five components described, which represent a computational model that implements the conceptualization of the fundamental needs existing in applications, especially to perform entirely the process explained in Chapter 2 of embedding information and knowledge about group activity into CSCL applications in an efficient and effective manner (see Figure 3.9). Indeed, these areas are characterized by a high degree of user interaction. This interaction generates many events which will be handled and analyzed by the *CSCL Knowledge Management* component so as to extract useful knowledge from the group activity and then to dynamically notify this knowledge about the activities of others to participants by means of the *CSCL Functionality* component. This is wholly performed by the two components related to the knowledge management and functionality support, namely *CSCL Knowledge Management* and *CSCL Functionality*, which are explain in great detail here.

3.3.2 CSCL Knowledge component

The CSCL Knowledge Management component (Caballé et al., 2007e) will manage and analyze all the specific and large user events in order to record user interaction data as information which is crucial for the correct control and administration of the collaborative learning applications. Therefore, this component completely specifies and implements

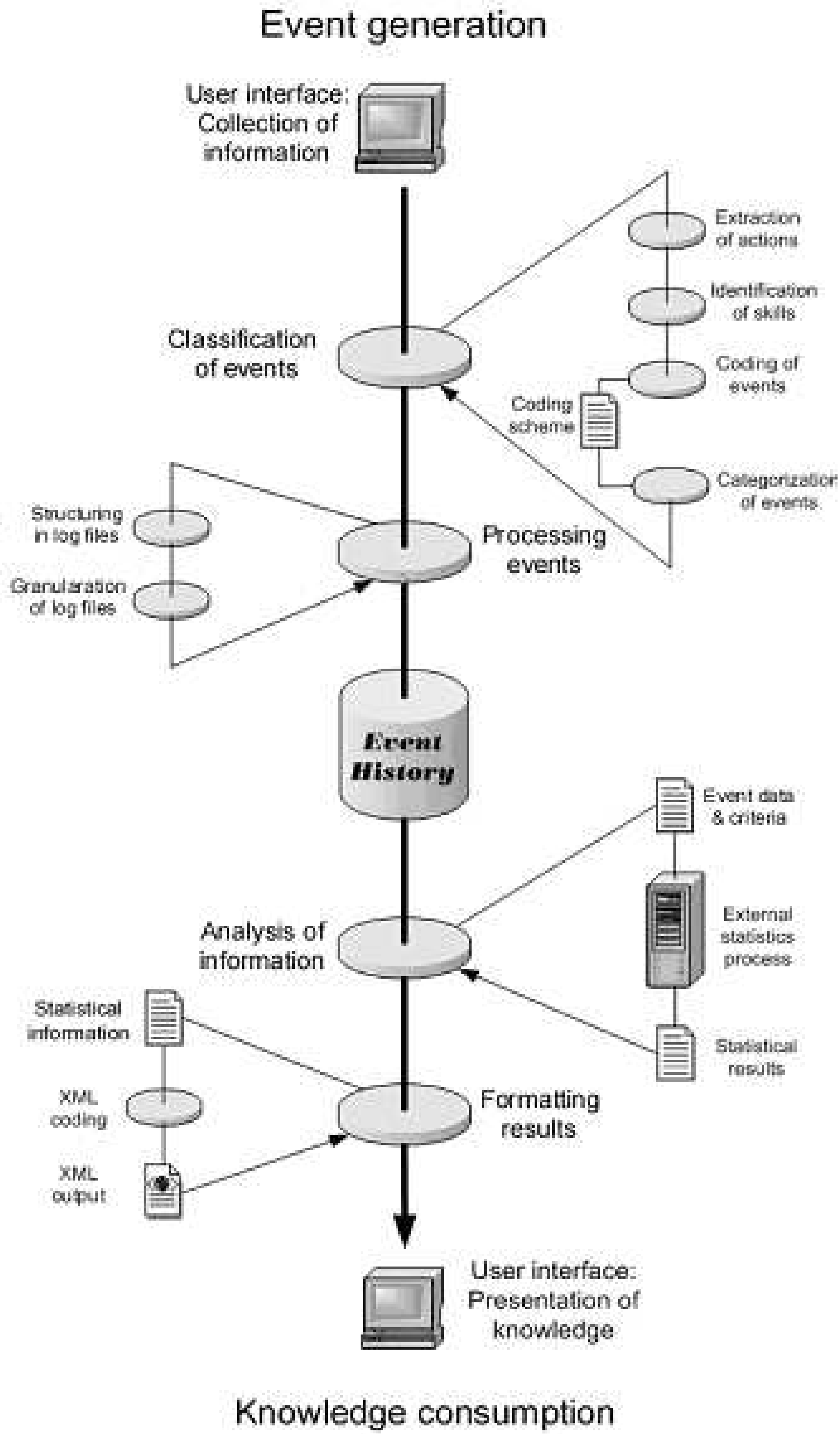

Figure 3.9: The process of embedding information and knowledge into CSCL applications. This picture is taken from Chapter 2 and reproduced here again for greater convenience

the first two stages (collection of information and analysis) of the process of embedding information and knowledge into CSCL applications (see Figure 3.9).

The final objective of this component is to extract valuable information from the events processed for later statistical analysis with the aim of revealing useful knowledge from the group activity. This component is made up of the two following subsystems ((see Figure 3.10)):

- *CSCL Activity Management* subsystem. This manages the system log files made up of all the events occurring in the system during a period of time. It represents the collection and classification of the event information as a source of information for the creation of the pertinent statistics.
- *CSCL Knowledge Processing* subsystem. This performs the management and maintenance of statistical analysis through the generated events stored in log files. It represents the analyses of the information previously obtained thus providing the necessary knowledge to control and facilitate the collaborative learning process as well as improving the general performance of the system.

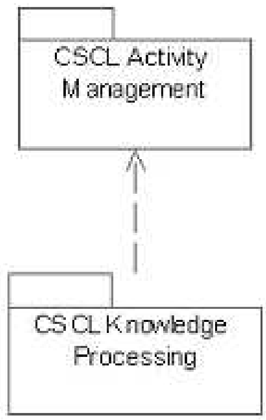

Figure 3.10: Graphical representation of the subsystems making up the CSCL Knowledge Management component.

Due to their importance, both subsystems are now described in more detail:

3.3.2.1 CSCL Activity Management subsystem

This subsystem (Caballé, et al, 2007e) collects, classifies and structures the event information contained in the CSCL application log files so as to make it possible to facilitate its later statistical analysis.

The *log file* is a key entity made up of all the action events occurring in the system over a given period of time and is automatically generated by the system during its usual functioning. This represents the source of information that is later used for the creation of the appropriate statistics. As mentioned in Chapter 2, in CSCL applications there is a need for the classification of all types of user and system events generated according to the three generic group activity parameters (see Chapter 2, Section 2.1 for the rest of this sub section). To this end, a complete and tight hierarchy of events (see Figure 3.11) is provided in this subsystem. In this hierarchy, a certain degree of redundancy is allowed

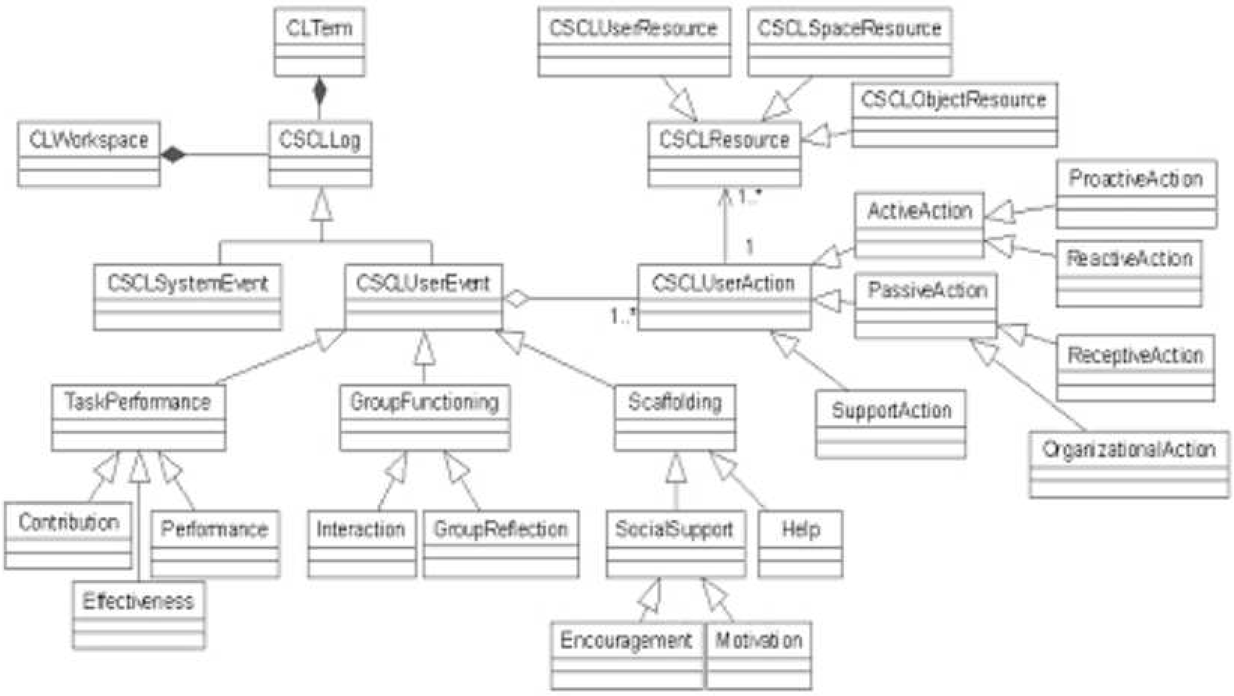

Figure 3.11: A class diagram to collect and classify all events generated during the group activity and for each academic term.

since both the same events to measure different elements are expected and desired (e.g. a group processing event can be simultaneously addressed as both a quantitative parameter to measure group functioning and a qualitative parameter to measure scaffolding). Furthermore, the group activity is driven by participants' actions on the generic collaborative learning resources and these actions are aggregated to the user events to form another hierarchical tree (see Figure 3.10) in which, at a first level, it is distinguished between active and passive user actions depending on whether or not the student contributes directly to achieving the group objective. At this same level, the support action (i.e. help, motivation and encouragement) is also considered and constitutes another distinct category.

In order to correctly classify the user actions on the resources during group activity according to the event hierarchy, this subsystem provides a classification process which includes a coding scheme (see Table 2.4 in Chapter 2) based on the conceptual model described in Chapter 2 for data analysis and management in asynchronous environments. In this process, the event information collected from the log files is handled in sequential steps consisting of extraction, identification, coding, and categorization (Caballé et al., 2005a).

Thus, firstly, the specific action performed by a user on a resource is extracted from the log files. Secondly, this action is interpreted depending on the type of event that was involved. This represents the essential information in the identification of the real intentions or skills shown by the user (e.g. creating a note during a debate can be interpreted as either revision or reinforcement of the information depending on whether the note was created in the context of a reply or as an observation). Then, during the third step of the process, the user event is uniquely codified according to both the user action performed and the real user skill identified in the context of the action. Finally, the user event is categorized into one of the above-mentioned group activity parameters according to the coding scheme shown in (see Table 2.4).

Given that this classification process is highly generic, it is only provided the most abstract form of categorization based on the above-mentioned generic event hierarchy (see Figure 3.11). Thus, the specific applications using this process should categorize their event information according to their particularization of this categorization.

Although it is possible to use the same classification process for both synchronous and asynchronous environments, this subsystem focuses on the latter as this is still the most usual way to collaborate in on-line collaborative learning environments and permits the complete automation of the classification process. Even so, recently, synchronous applications are receiving a lot of attention since they allow for the extraction of knowledge from real-time interactions from which it is possible to understand essential aspects of on-line collaborative learning such as socio-cognitive approaches to culture, social construction of community and so on. To this end, this subsystem is flexible enough to handle information from synchronous environments by incorporating another coding scheme and codifying the user actions manually.

Once the event information generated in the group activity is correctly collected and classified, this subsystem structures this information in such a way that it is possible to both prepare information to facilitate its later processing and analysis and allow it to be addressed in a distributed environment where available (such as in a Grid environment (Foster and Kesselman, 1998).

To this end, the following generic steps are proposed so as to correctly structure the event information for later processing (see Figure 3.9 and Caballé et al., 2007f): on classifying the event information and turning it into persistent data, this information is stored in the system as structured files, which will contain all the information previously collected from the system log files in specified fields. Next, the files are structured in accordance with two basic criteria, time and workspace, which characterize all group collaboration. These files will represent as high a degree of granularity as possible for both criteria and they will be parameterized so that the administrator can set them up in accordance with the specific analysis needs.

Thus, every single workspace will have its own structured file made up of all the events occurring within it for a given period of time and, hence, during later data processing it will be possible to concatenate several structured files so as to obtain the appropriate degree of granularity (e.g. all the groups in a classroom for each 12 hours) and so making it possible to efficiently parallelize the data processing depending on the characteristics of the computational resources.

Therefore, in the context of distributed systems, CSCL applications can take great advantage of the parallelism to process several files (e.g. all the groups in a classroom) at the same time and thus dramatically reduce the overall computational time. As a result, it is possible for these applications to process a large volume of collaboration activity data and make the extracted information available in real time.

At this point in the process of embedding information and knowledge into CSCL applications, the information has been collected, classified and well-structured so that it can easily and efficiently be processed and analyzed during the work of the *CSCL Knowledge Processing* subsystem.

3.3.2.2 CSCL Knowledge Processing subsystem

After the event information from the structured files has been processed, the results of data processing are stored in a database manager system, in which all the information contained in the structured files should be correctly represented, even if they are distributed in different machines. The aim is to make it possible to consult both the desired data from the database directly (e.g. the number of connected users, the type of documents in a certain workspace) and the computed complex statistical results produced from processing these data (see Figure 3.9).

Therefore, it is necessary to design the database correctly so as much useful information as possible can be stored and computed and the desired information can be efficiently extracted. Thus, based on the premises argued in (Watson, 2003), the design of the database should be generic, reliable, reusable, efficient and independent from any specific database manager. To this end, a proposal of the design of the database is provided[3] (see Figure 3.12) for the conceptual design) which satisfies all these requirements and allows users to consult data regarding the basic entities that take place in any CSCL application (users, objects, workspaces, connections, etc.).

The last step in this subsystem (Caballé et al., 2007e) is to use the resulting information in the databases to compute statistical results and present them to the members of the online collaborative group and the tutors through the CSCL Functionality component (see Figure 3.5). In this context, an XML coding of the statistical results is suggested in order to make it possible to present this information to final users in different forms (see the stage of formatting results in Figure 3.9). Considering the highly structured nature of the data and the design of the relational database (see Figure 3.12 and Fernández et al., 2002), this last step should be designed as a middleware (Kyung-Soo, 2001) which performs the following functions: to extract necessary information from the databases, to compute statistical measurements as desired, and to convert the results into XML output. This design will provide sufficient flexibility as to allow ad hoc statistical measurements to be obtained as well as permitting the creation of user-specified document type definitions (DTD) to accommodate the different needs of information representation.

To sum up, the *CSCL Knowledge Processing* subsystem carries out the statistical analyses and interpretations of the event information generated during group activity which has been previously collected, classified, structured and stored in a database. To this end, a statistics module will have the information source stored in the database along with generic criteria with the aim of performing the desired quantitative analysis of the individual and group collaborative activity and learning. To make it possible, a data structure of generic and parameterized criteria (see Figure 3.13) is designed to classify the most usual requests for information in CSCL environments (e.g. ”How many users accessed the system during a period of time?”, ”Which users read a document?”) thus making it possible to reuse them in as many statistics as possible. Furthermore, a statistics abstraction is provided with the potential to extract useful data from the database which in turn will give detailed knowledge of those parts of the system (e.g. workspaces) or the users’ activity interested.

These statistical results should be obtained by the application server as fast as possible and presented to group members and tutors preferably in real time and in different formats if necessary. Both immediate user awareness of what is going

[3]See the complete design of the database at: http://clpl.uoc.edu/docs/DBDesign.pdf (Web page as of April 2008)

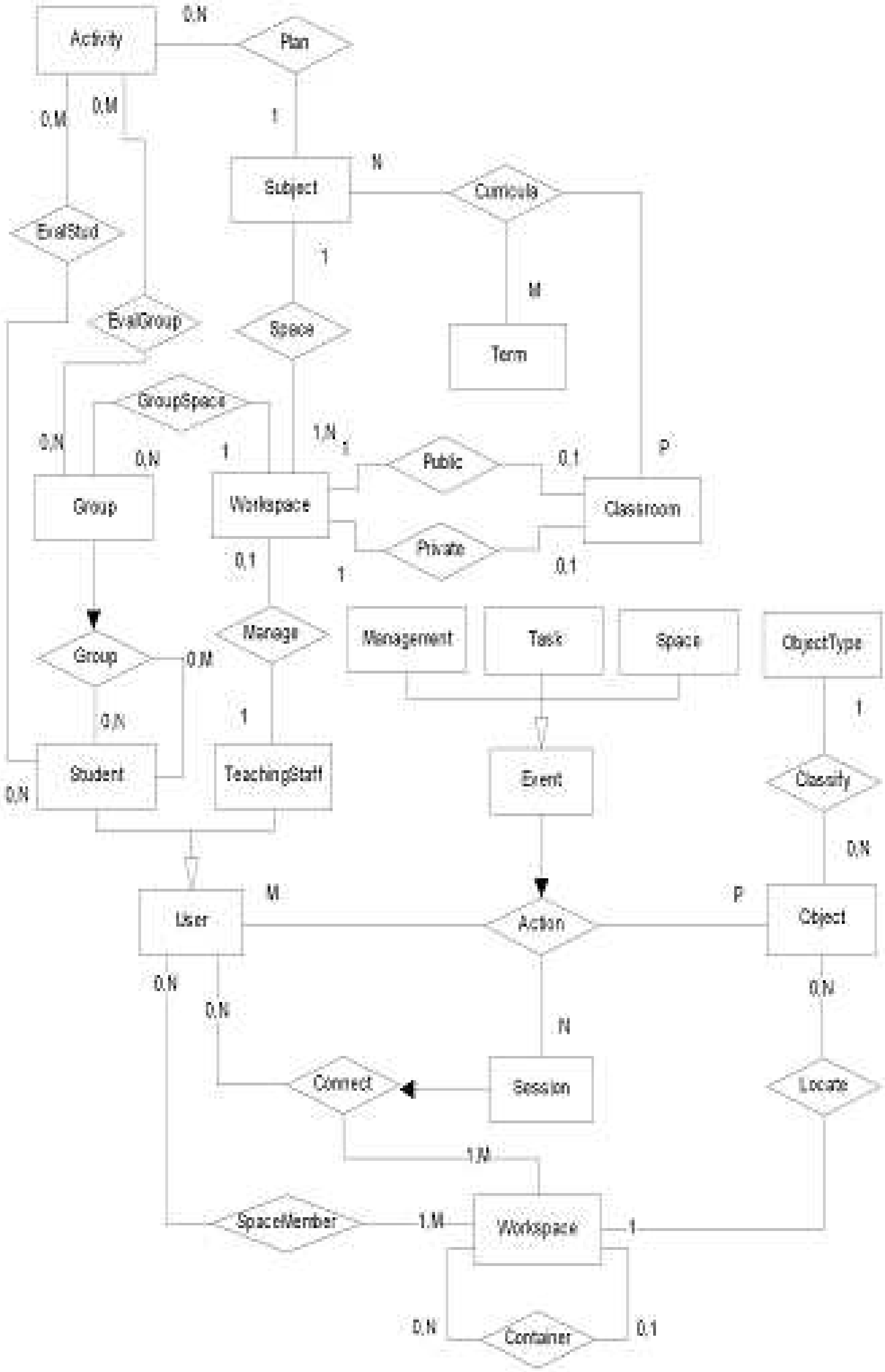

Figure 3.12: ER diagram representing the conceptual design of the database.

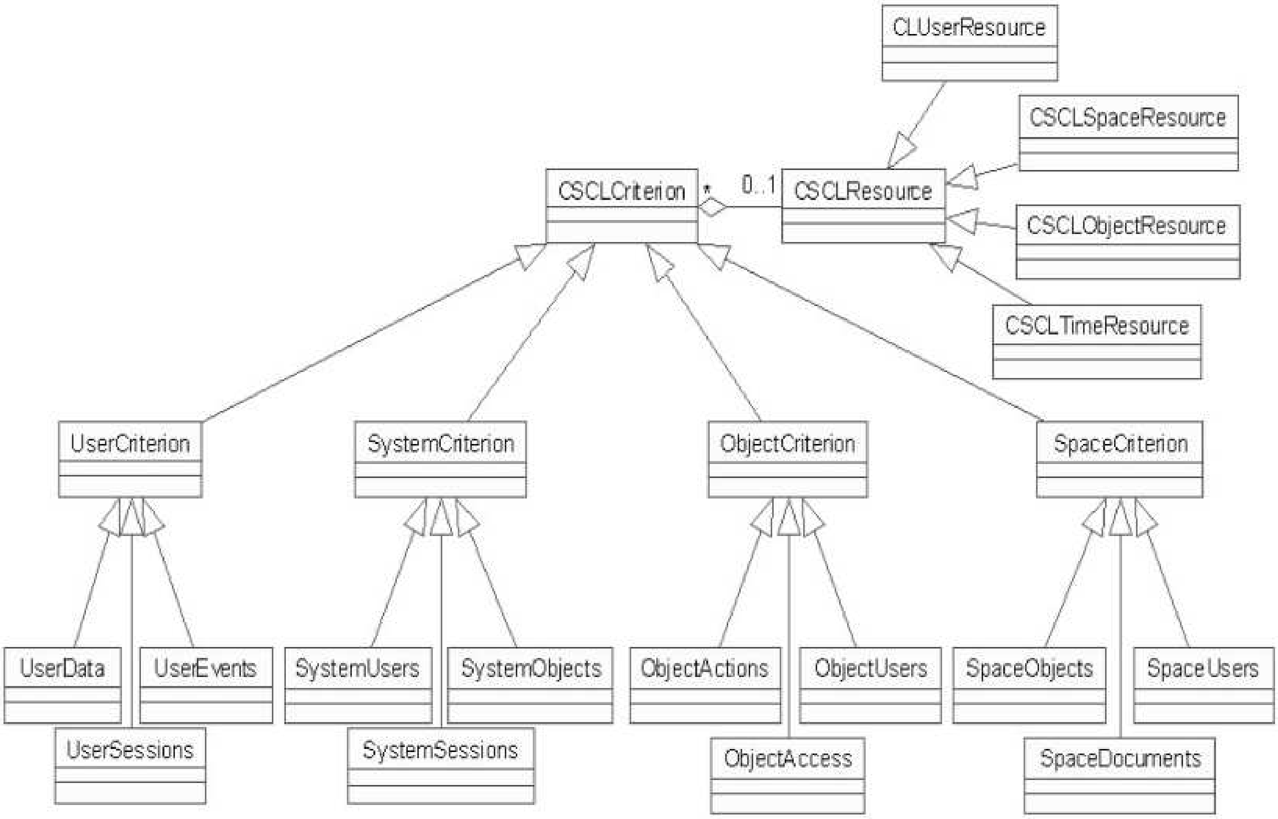

Figure 3.13: A hierarchy to classify criteria in CSCL applications.

on in the system and constant feedback on group and individual performance are achieved by continuously processing the current events. To this end, this subsystem is intended to exploit the most powerful statistical packages in the market through the creation of a generic interface to make it possible both to export data to the most popular statistical formats and to import the obtained results.

The ultimate objective of this subsystem is to define a bottom-up analysis approach that processes and analyses the user events in order to decode the specific actions of the users describing their interaction during the collaboration activities. The analysis aims at identifying those sequences of actions that can be used to determine typical patterns of interactions (Inaba et al., 2003). Thus, at this point in the research the objective is to identify as many best CSCL collaborative learning practices as possible, which can then be translated into typical collaborative learning patterns.

Based on a model of desired interaction, the system allows tutors to compare the learners' real interaction processes with the typical interaction patterns in order to infer whether or not the process is effective for the learner. Furthermore, the knowledge revealed by this analysis can enhance self and peer evaluation, which in turn improves the efficiency of group activities, monitoring group behavior and the individual attitudes of its members in the shared workspace. In addition, this knowledge is useful in assisting the tutor by providing the necessary means to support and assess individual and group learning outcomes.

3.3.3 CSCL Functionality component

This component (Caballé et al., 2007e) forms, along with the *CSCL Knowledge Management* component, the basis of the collaborative learning applications by defining the above-mentioned three elemental parts involved in any CSCL collaborative learning application: coordination, communication and collaboration (Ochoa et al., 2002; Caballé et al., 2004). Coordination involves the organization of groups to accomplish the important objectives of members such as workspace organization and group structure and planning as well as allowing group monitoring, which is achieved by maintaining the awareness of its participants. Collaboration lets group members share any kind of resources in both synchronous and asynchronous modes. Finally, communication represents the basis of the whole component since it enables coordination and collaboration to be achieved (Ochoa et al., 2002) by providing them with low-level communication support (basically by means of both synchronous and asynchronous interaction). Moreover, communication is also involved in high level interaction support during a message exchange session (email), which can be performed within and between groups. In a CSCL application, these three aspects will generate many events that will be thoroughly processed and analyzed (through the CSCL Knowledge Management component) to extract knowledge that will be provided to the users as required.

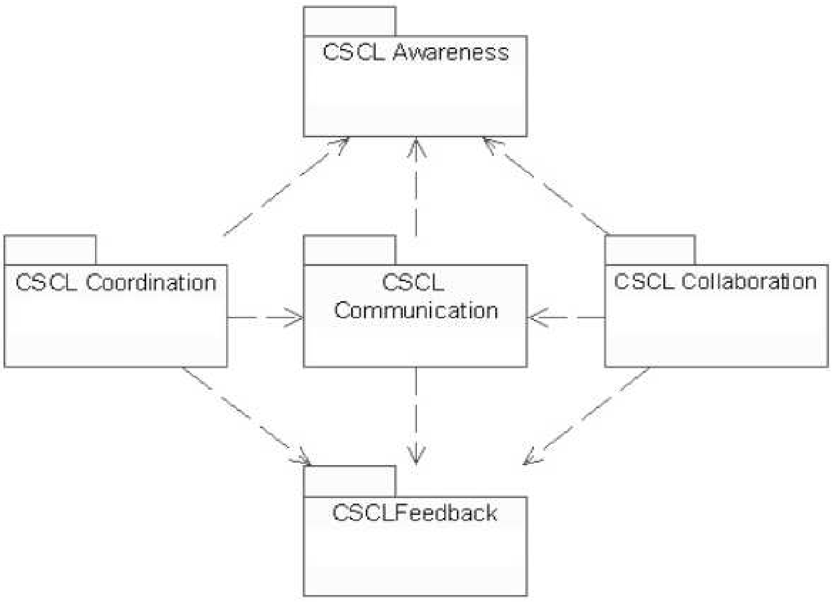

Figure 3.14: Graphical representation of the subsystems making up the CSCL Functionality component.

The final objective of this component is to provide functional support to CSCL applications in terms of coordination, collaboration and communication (see Figure 3.14. Moreover, this component implements the last stage of the process (see the stage of presentation of knowledge in Figure 3.9) of embedding information and knowledge into CSCL applications (i.e. presentation of the knowledge generated) by providing the users with immediate awareness and constant feedback of what is going on in the system. This stage is mainly managed by two specific modules (CSCL Awareness and *CSCL Feedback*) of the component, which has five subsystems in all (see Figure 3.15 and Caballé et al., 2007e):

- *CSCL Coordination.* This manages both members and resources within a collaborative group so as to make it

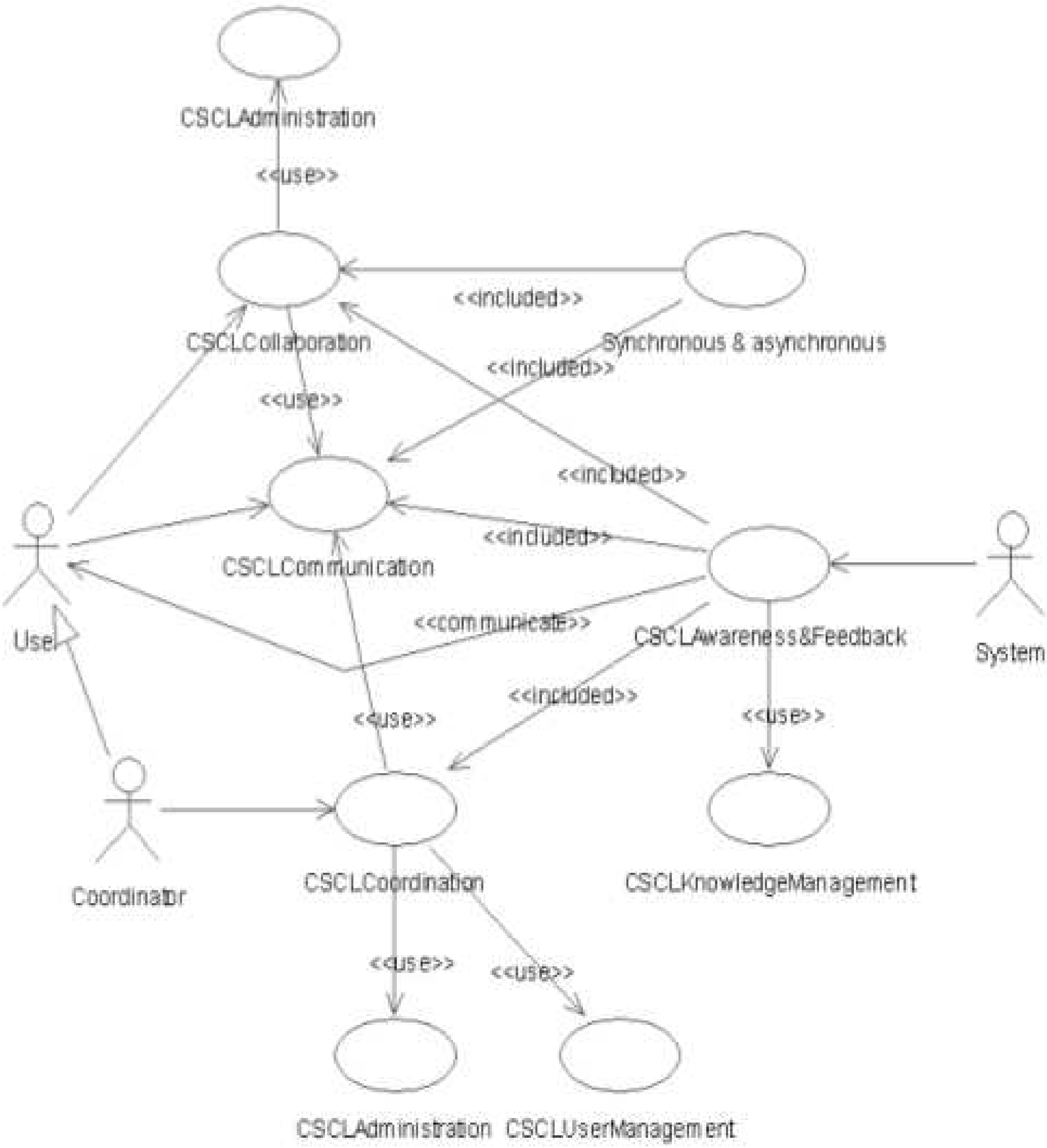

Figure 3.15: The analysis of the CSCL Functionality component.

possible for members to accomplish their objectives.

- *CSCL Collaboration.* The main purpose of this subsystem is to let participants share resources such as files and applications in a collaborative learning environment. Resource sharing may be in both synchronous and asynchronous modes.

- *CSCL Communication.* This manages all the low-level interactions between two or more participants within a collaborative learning group in both synchronous and asynchronous modes.

- *CSCL Awareness.* All the events generated by the previous subsystems during a session will be captured and managed by the CSCL Knowledge Management component and will form the awareness information which will be immediately notified to system users.

- *CSCL Feedback.* This aims to influence group participants in a positive manner by means of a steady tracking of parameters outside the task itself and by giving constant feedback of the information related to these parameters to the group.

3.3.3.1 CSCL Coordination subsystem

Coordination (Caballé et al., 2007e) is an important aspect of any collaborative activity. It entails the combination and sequencing of otherwise independent work toward the accomplishment of a larger goal.

In collaborative learning environments, coordination mostly refers to the tasks that lead toward the learning group formation as well as the definition and planning of the group objectives. Moreover, the group coordinator may track task status, deadlines, resource usage, working results, or other critical process parameters so as to lead the group correctly.

The coordinator (as the facilitator of the group), who may be an independent member such as a teacher, incorporates participants and resources in the group and assigns them to the well-defined tasks and activities for the performance of the planned project. Once the group has been organized, there will be continuous monitoring and control of the tasks accomplished, group member participation, resource usage, project deadlines, course results, etc. In order to perform the control tasks, it is necessary for the coordinator to be aware of all the events generated by group members so as to make up the appropriate reports of the group activity.

This subsystem aims to provide the basic functionality to organize and coordinate the collaborative learning group for CSCL applications. To this end, some fundamental abstract entities, such as an *appointment*, making up others such as *calendar*, *agenda*, *notice board* and so on, as well as the basic operations to manage them, have been created. These generic entities provide such information as course description, evaluation policy, course schedule, links to study resources and the most up-to-date information sent to the participants from the coordinator. The aim is to provide full support of the group organization and planning as well as task management.

In order to make it possible for the coordinator to both monitor and assess the group activity, a generic report system is provided based on the entity *report* with the aim of keeping track of the performance of participants. This includes, amongst others, scoring for both individual and group-supported task evaluation purposes as well as reporting of group results at the end of the collaboration into basic activity parameters such as task performance and group functioning.

Furthermore, this subsystem provides a basic tool to facilitate the coordination of the group in the form of a generic multi-parameterized search function that enables participants and resources to be easily located by different criteria. As for all basic communication support that is necessary to perform the usual coordination activities (e.g. making an appointment in the agenda, giving news to participants, etc.) is provided by the *CSCL Communication* subsystem, as seen later on. Finally, information about group members and resource usage will be provided by the *CSCL Users Management* and the *CSCL Administration* components respectively (see Figure 3.14), where these data should previously have been handled.

3.3.3.2 CSCL Collaboration subsystem

Collaboration (Caballé et al., 2007e) relies on participants sharing all kind of resources. In a sharing Web-based environment, students, as participants in a group, share resources amongst themselves in order to achieve their learning objectives. It should be noted that sharing resources, in the context of this research, means those applications that are shared by the participants (examples of these applications might be chats, multi-user editors, whiteboards, document repositories and so on). To this end, each participant collaborates by incorporating their contributions into the sharing resource so as to elaborate the problem by proposing solutions, discussing the contributions of others and so on. The sharing of resources amongst several participants is therefore a central functionality of collaborative learning applications.

In this context, *floor control* (El Saddik et al., 2001) is essential to the proper functioning of collaborative sessions as this avoids concurrency problems by allowing only one participant at a time to control the shared application. Without floor control, there would be collisions of events, which would have undesired consequences (Caballé, at al, 2007e).

The coordinator is the only group member or independent user with the necessary privileges to start and close a sharing session. The rest of the users (the other group participants and the coordinator when acting as a group member) can only consult and modify the sharing resources (e.g. making a new contribution to a shared application such as a multi-user editor resulting in a modification to the state of the application).

Sharing between different users may be synchronous, with different participants accessing the same resource at the same time (seeing and working on the same copy of the resource such as multi-user editors), and asynchronous, with several users accessing the same resource at different times (each of them working on a different copy of the same document with the possibility of changes of participants in the shared resources being flagged to the others).

This subsystem aims to provide the basic functionality to share any kind of resource within the collaborative group activity. To this end, the most fundamental abstraction here is a *room*, which is defined in this subsystem to represent all virtual spaces (i.e. sharing applications) where any kind of sharing is performed. Thus, in a room it is possible for participants to share a conversation (i.e. chat), a text or graphic content (i.e. multi-user editors) or a pool of study material (i.e. document repository) among others activities. The result of the interaction is then updated in the sharing application to other participants by the mechanisms provided by the *CSCL Communication* subsystem, as seen later on.

In this subsystem, basic utilities for improving the collaboration such as floor control and session moderation among others are also provided. Thus, a generic artifact called *semaphore* is provided for concurrency or floor control purposes during synchronous collaboration (e.g. multi-user editor). This artifact allows a participant to properly interact with the shared application. The subsystem first restricts the access to the shared application by locking a semaphore denying access to other participants trying to interact with the application. When the first participant is finished, the system releases the semaphore and others can take control of the application.

In a collaboration context it is also important to take into account session moderation in order for a collaborative session to be more productive. In the CLPL these needs are met by using the coordinator features of the *CSCL Coordination* subsystem. All the events generated during sharing will be handled by the *CSCL Awareness* subsystem, which will immediately notify users of the current state of the application (e.g. a new contribution, deletion of a document of the repository, etc.). This subsystem will in turn use the *CSCL Knowledge* Management component (see Fig. 9) where these events will have been previously handled. Finally, low-level communication support, which is necessary in the performance of all sharing activities (e.g. making a contribution to a news group, participating in a multi-user editor session, etc), is provided by the CSCL Communication subsystem, as seen later on.

3.3.3.3 CSCL Communication subsystem

Communication (Caballé et al., 2007e) is the most important functional aspect in groupware systems aiming to support the communication between two or more collaborative learning participants. Furthermore, in any CSCL application, coordination and collaboration rely on communication to accomplish their objectives. Communication includes text messages, spoken interactions, or non-verbal exchanges like gestures in a video conference (Baloian et al., 2004). The communication support is based on four elements: a *message* as the information carrier between a *sender* process and a *recipient* process, which receives and possibly processes the message through a *channel* (Ochoa et al., 2002). Moreover, in this context, it is necessary to implement different ways of message communication such as point-to-point, multicast and broadcast. This subsystem manages an interaction between two or more users (i.e. student, lecturer, coordinator, administrator, devices, etc) in a CSCL application.

Communication may take place synchronously or asynchronously. In a synchronous mode, interaction takes place when participants communicate between each other at the same time such as during a chat or video conference session. This type of interaction also refers to a piece of information (e.g. a line of chat, a spoken contribution to a video-conference, or a paragraph in a shared text editor) that is delivered by a user and reaches its peers at the same time. Although this kind of interaction is usually destroyed once the sharing session comes to an end, in the context of this research, it will be deliberately stored in log files once the sharing environment is completed. This way, it is possible to classify, structure and analyze this information as a part of the process of extracting knowledge from communication, which is a low level of user interaction.

In an asynchronous mode, on the other hand, the interaction takes place when different participants communicate at different times and it especially refers to communication involving participants interacting with others employing different asynchronous tools (e.g. e-mail, news, forums, debates, stick-on notes). This kind of communication goes beyond a simple conversation and so it must be stored in the system as an important source of information about group activity.

It is also possible to have interaction types which, like the instant messenger tool, employ both synchronous and asynchronous modes. The main purpose of this subsystem is to receive and manage all low-level interaction generated by group activity. During coordination (e.g. showing a warning to all group members) and collaboration (e.g. propagating a new line of chat to the session participants) and in both synchronous and asynchronous modes, communication will act as a channel to transport the low-level content of the interaction (e.g. the text of a warning message, the phrase of a line of chat, etc.) and propagate it to participants if necessary.

Furthermore, this subsystem is responsible for providing support of certain types of high-level interaction between participants in which neither coordination nor collaboration are involved (e.g. sending a single email). Therefore, in this context, this subsystem performs all of the work managing both high- and low-level interaction.

From a generic view, communication permits any kind of textual or non-textual exchange such as an email transaction or a multi-user editor session between a sender and a recipient. To make the communication possible, it is necessary for the message to be sent to a recipient, which forwards it to the addressee.

This subsystem abstracts the network complexities and the specific communication needs by defining four fundamental generic entities, namely message, sender, recipient and channel, which form, as seen before, the basic communication infrastructure of any collaborative system. At this level of genericity, the specific CSCL applications are expected to particularize these four abstractions into the specific underlying communication infrastructure (channel abstraction), which can be centralized, be fully-distributed, or be a hybrid of both. Thus, this subsystem can be implemented in many different ways to meet specific communication needs.

Therefore, as an example of synchronous communication, a line (message) drawn by a participant (sender) during a whiteboard session would be sent to the whiteboard communication module (recipient) where it would then be immediately propagated to all other participants so as to update their user interface. When the communication takes place asynchronously, such as in an email transaction, the sender forwards a message to the addressee through the mail server (recipient), which usually stores it before being received by the addressee.

All interaction between users generates events (e.g. a user gets an email, a participant receives somebody else's line in a chat room, etc.). All these events will be collected and managed by the *CSCL Knowledge Management* component (see Figure 3.15 and will form the awareness information which is immediately presented to system users by the *CSCL Awareness* subsystem making others aware of what is going on in the system and, above all, in the learning group's workspace during the group activity. Furthermore, awareness is essential in this context since it makes it possible to show the availability of other participants for communication.

It must be mentioned here that there is expected to be a certain redundancy when different subsystems generate events while involved with the same resources at the same time. It should be noted, however, that the communication module, when acting as a transport mechanism, generates lower level events than those generated by other modules. Thus, as an example, a new line of chat during a chat session may be considered as being both collaboration in the modification of the shared resource (i.e. the chat room) and as communication in the forwarding of the message to a recipient (i.e. chat communication module). Therefore, both views are complementary and the *CSCL Awareness* will unify them to present more complete information to users. This is described in more detail below.

3.3.3.4 CSCL Awareness subsystem

Awareness (Gutwin et al., 1995; Caballé et al., 2005a) is essential for any of the three forms of cooperation seen. It allows implicit coordination of collaborative learning, opportunities for informal, spontaneous communication, and it keeps users informed as to what is happening in the system (Caballé et al., 2004). When awareness is synchronous, users know in real time what other co-participants are doing (e.g. during a multi-user editor session, who is editing and what is being shown) and which documents are being used by others whereas when it is asynchronous, users receive delayed knowledge of who, when, how and where shared resources have been created, changed or read by other users.

Users are continuously interacting with the system (creating documents, reading emails, etc.) and so events are generated which, once they have been collected, classified and processed, will then be gathered, managed and formatted by the system as awareness information for distribution to the users involved. This awareness process will use the *CSCL Knowledge Management* component (see Figure 3.16 where the new events will previously have been managed before communicating them to the users. This process is triggered by the system automatically and so is transparent to non-participating users.

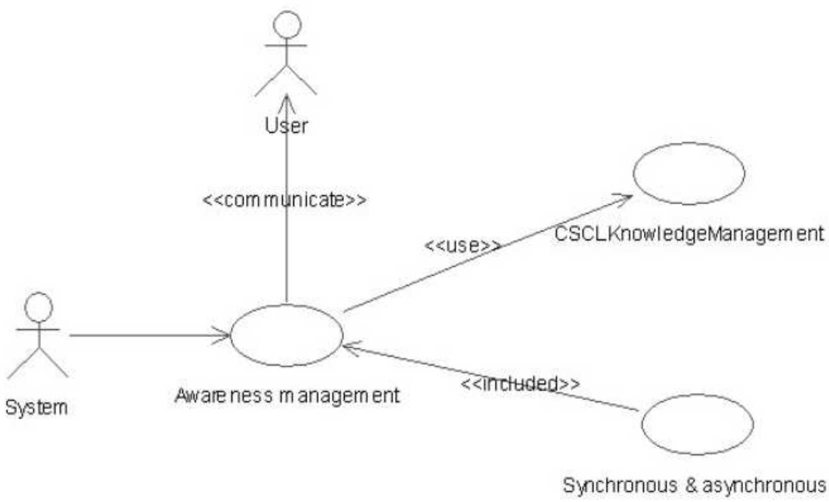

Figure 3.16: The analysis of the CSCL Awareness subsystem.

Therefore, when users communicate and collaborate with each other and also when they are coordinating their activity, they must always be aware of both what other users are doing at the same time and what other users did in the past. This implies receiving information from both synchronous and asynchronous modes (see Figure 3.16 and being aware of both the resources used and the users who interact with the system.

In a synchronous context, the user should be aware of the current activity in the group (the contribution of other members, their location and availability, the users working on a shared document at the same time and so on) and this information should be provided by visual and audio signals. To develop efficient synchronous in CSCL applications providing transparent awareness to users who may start at both the same and at different points in time is a significant challenge. Group activity events which have been collected and analyzed (including the transformation of information into a serializable structure) have to be distributed to all other participants. The distribution may be performed by a central dispatch server, where session participants are registered, or by using a multicast-capable publish/subscribe communication infrastructure.

In an asynchronous context, on the other hand, users need to be aware of the activities performed by others as a basic requirement to solving the task at hand and supporting asynchronous communication (e.g. flagging incoming mails).

Event redundancy is an essential property during group activity in CSCL applications and, as seen before, it is desired and expected to ensure the provision of users with complete and effective awareness information. Thus, during any interaction for coordination, collaboration, and high-level communication purposes, low-level events from the communication transport mechanism will be generated on certain resources at the same time as other subsystem involved are generating high-level events on the same resources. As a result, there will be more event information to be managed and thus the resulting awareness information will be more complete. As an example, in making a new contribution to a news group (collaboration purposes), those low-level events generated by the communication subsystem will only inform about the sender, the recipient (news group server) and the content of the message sent.

In the case of high-level events generated by the collaboration subsystem, they will contain information about the type of contribution (e.g. new, reply, help, etc.), the specific news group to which the contribution has been made, the topic of the news group, the date and time and so on. Therefore, by unifying these two types of events it is possible to obtain complete awareness information about the contribution made.

In order to provide the essential awareness information to effectively support the three areas seen, this subsystem (Caballé et al., 2007e) defines three generic entities, namely *resource state*, *user status* and *group memory*, which support the collaboration, communication and coordination respectively. Each of these abstractions acts as a vehicle so that awareness information can be classified and presented to users in the correct form depending on the type of activity involved.

Therefore, first, in sharing (e.g. a multi-user editor session), participants are continuously modifying the state of the shared application (e.g. writing a new text comment, deleting somebody else's sketch, etc.) and so the current application *state* has to be continuously propagated to users as a news warning signal. Secondly, it is essential to show the current participants' *status* so as to be aware of the availability of them for communication (e.g. before sending a message to others it is crucial to know whether or not they are available). It should be noted that availability refers to both synchronous (online availability during a chat session) and asynchronous (the participant who an email is addressed to must be registered in the system). Finally, the persistent storage of awareness information as *group memory* (Conklin,

1992) is essential during coordination since it allows for document and data access, which are commonly stored for later retrieval, and also the context in which they were created. Thus, being aware of the activities of others is essential for coordination (e.g. decision-making, group organization, social engagement, etc.).

Furthermore, for the purposes of presentation format, this subsystem defines a *flag* as a single abstraction supporting the presentation of awareness information to users through the user interface by any means: from a visual and simple sign for warning purposes (e.g. a new participant has just joined the chat) to complex visual and audio effects to keep participants aware of what is happening in the group activity. It may also include different types of short text information provided to report the most recent event history on a specific resource as well as certain statistical analysis results of the group activity.

The ultimate objective of this subsystem is to present awareness information to users in a correct, effective and immediate fashion as the last stage in the process of embedding knowledge into CSCL applications (see Figure 3.9).

3.3.3.5 CSCL Feedback subsystem

Feedback (Zumbach et al., 2003; Caballé et al., 2005b) in collaborative learning environments is receiving a lot of attention due to its positive impact on the motivation, emotional state, and problem-solving abilities of groups in on-line collaborative learning (Caballé et al., 2005b). It aims to influence group participants in a positive manner by means of a steady tracking of parameters outside the task itself (such as motivation and emotional state) and by giving a constant feedback of these parameters to the group. Therefore, when users participate in a application, they may enhance their abilities by increasing their knowledge about others in terms of motivation, interaction behavior and so on.

Feedback goes one step further than awareness by providing exhaustive information of what is going on in the group over a long period of time (e.g. constantly showing to each group member the absolute or relative amount of the contributions of others). Furthermore, feedback may be obtained about the emotions and motivation of participants through asking them about these states. In all cases, feedback implies receiving information simultaneously both synchronously and asynchronously since the history information shown is continuously updated.

During the feedback process, all new information communicated to the users will have been previously collected, classified and analyzed by the CSCL Knowledge Management component. As a consequence of the complex knowledge provided to participants in form of feedback (e.g. group's member relative and absolute amount of contributions, group's members variation in motivation and emotional state during last two hours, etc) this subsystem makes a strong use of the statistical analysis and need to show the results obtained in complex graphical formats.

In this subsystem Caballé et al., 2007e) certain generic entities are defined such as *history, pool* and *diagram* and functions such as sorting. Based on these abstractions it is possible to dynamically gather and store great amounts of history data and statistical results from the group activity in order to constantly update and present them to participants in the appropriate diagrammatic form (e.g. pie chart, histograms, etc.).

So far, the Platform Independent Model have been described to model a generic, reusable approach of the CSCL domain. Next section deals with the provision of technology to the PIM model

3.4 Software technology for systematically engineering CSCL applications

Following the principles for GP and MDA developments, once these five components forming the PIM of the CLPL have been fully analyzed and designed, they are to be realized using specific technologies. To this end, two different Platform Specific Model (PSM) have been constructed so far from the PIM: An Object-Oriented (OO) approach by means of a Java implementation and an approach that follows the Service Oriented Architecture (SOA) principles. Both technology approaches are described in detail and justified their use for the realization of the CLPL, especially from the GP standpoint.

3.4.1 The feasibility of Java for the construction of reusable CSCL software

The first PSM of the CLPL with Java programming language due to its great predisposition to the adaptation and correct transmission of generic software design, which make the software highly reusable (see Caballé and Xhafa, 2003a and an example in Figure 3.17). To this end, in order to encourage the reusability of the CLPL components the basic requirements forming the PIM are designed separately with OO methodology. In order to maintain intact the ideas of GP design that are found, an implicit logical layer is implemented that creates a correspondence between the GP and OO design.

```java
package clpl.users;

import gpl.users.*;
/**
 *  Abstract class carrying out the logic part of the group creation within a CSCL environment.<br>
 *  @author <a href="mailto:scaballe@uoc.edu">Santi Caballé Llobet</a><br>
 *  @see <a href="http://cv.uoc.edu/~scaballe/gpl/api/gpl/users/GenericUserCreation/">gpl.users.GenericUserCreation</a><br>
 *  @version 1.0a; 15-12-2003; java version "1.2.2"<br>
 */
public abstract class CLGroupCreation extends GenericUserCreation{

    /**
     *  Constructor without parameters from the class CLGroupCreation.<br>
     *  Creates a group.<br>
     *  (pre: the group does not exist in the data store.)<br>
     *  (post: the group has been created in the data store.)<br>
     */
    public CLGroupCreation() throws GenericUserException, Exception {
    }
}
```

Figure 3.17: An example of the Java-based PSM of CLPL as coded design.

In codifying the PIM of the CLPL in Java the main objectives of GP and Java's characteristics were matched:

- Reusability and extensibility allow software to adapt to many interrelated problems, which is the main aim of GP. Java has many mechanisms such as Object type and interface and abstract class which make the CLPL fully susceptible to reutilization. The independence of the platform makes this skeleton portable to most known environments.

- The great potential for the reutilization of GP makes it necessary to guarantee a level of maximum quality. Java has a powerful mechanism of exception management which increases the robustness of the library and hence its quality.

- The *javadoc* documentation provided by Java also increases quality by facilitating the test phase and maintenance. GP aims to create software which is as general as possible without losing efficiency by finding the most abstract form of software.
- The simplicity of Java allows the programmer to concentrate on the mechanics of specialisation without having to control minor details. Applications with strong user interaction, such as the library, minimize both the relative decrease in performance due to java being interpreted and the penalization for the casting on use of Object.
- The increase in productivity is obtained by the reutilization of existing components. Java has large stores of highly reusable useful code (data structures, etc.) that allows code to be written better and faster and so clearly favoring increased productivity. This Java-based PSM is faithful to this idea.
- Once generic software based on GP has been built, it is then necessary to personalize it to a subgroup of particular requirements so that a specific use within an iterative cycle of abstraction/personalization can be made of it. Due to Java's capacity, it is feasible to specialize the components of a generic library such as the CLPL components in different ways.

As a result this Java-based PSM is made up of five packages which constitute the skeleton of the basic structure of whatever application of this domain is constructed using this PSM. The Javadoc documentation of this PSM of the CLPL is found at http://clpl.uoc.edu.

As mentioned previously (see first section), the CLPL is a particularization of the General Purpose Llibrary (GPL) (Caballé and Xhafa, 2003b; Caballé and Xhafa, 2003c) platform, which was developed following the same principles of the CLPL. There exists also a Java-based PSM of the GPL whose Javadoc documentation in the form of Application Programming Interface can be found at http://cv.uoc.es/ scaballe/tfc/api.

3.4.2 On the advantages of using service-oriented architectures for CSCL

The second PSM of the CLPL was developed following the principles of Service-Oriented Architecture (SOA) and realized using Web-services (Caballé, 2008d; Caballé, 2007g).

SOA (W3C Working Group, 2004) represents the next step in the software development to help organizations meet their ever complex set of needs and challenges, especially in distributed systems. This is achieved by dynamically discovering and invoking the appropriate services to perform a request from heterogeneous environments, regardless of the details and differences of these environments. By making the service independent from the context, SOA provides software with important non-functional capabilities for distributed environments (such as scalability, heterogeneity and openness), and makes the integration processes much easier to achieve.

SOA relies on services. According to W3C Working Group (2004), a service is a set of actions that form a coherent whole from the point of view of service providers and service re-questers. In other words, services represent the behavior provided by a provider and used by any requesters based only on the interface contract. Within SOA, services:

- stress location transparency by allowing services to be implemented, replicated and moved to other machines without the requester's knowledge,
- enable dynamic access as services are located, bound and invoked at runtime,
- promote interoperability making it possible for different organizations supported by heterogeneous hardware and software platforms to share and use the same services,
- facilitate integration of other existing systems and thus protect previous investments (e.g. legacy assets),
- rely on encapsulation as they are independent from other services and their con-text,
- enhance flexibility by allowing services to be replaced without causing repercussions on the underlying systems involved,
- foster composition from other finer-grained services.

Organizations leveraging the key properties of SOA realize many benefits (Caballé, 2007g). By location transparency as well as dynamic discovering and invocation of a service, software mobility becomes a reality. This allows organizations to have the flexibility to move services to different machines without having repercussions on the underlying system involved. Furthermore, location transparency also promotes scalability and availability without the client's knowledge as it is possible both to scale the number of instances of multiple services which are running on multiple servers and to support fault-tolerance by redirecting a request to a replica when a server is unavailable.

Despite SOA can be realized with other technologies, over the last few years Web-services has come to play the major role in SOA by providing a set of standard protocols that meet the main needs of SOA, such as XML, UDDI, WSDL, SOAP and HTTP (for an overview see Chapter 1 and W3C Working Group, 2004). This is because these protocols allow a service to be platform - and language- independent, dynamically located and invocated, and interoperable cross over different organization networks. As a result, Web-services technology provides lower costs of integration along with flexibility and simplification of configuration.

There are a great deal of similarities between collaborative learning needs and benefits provided by SOA (Caballé, 2008d; Caballé, 2007g; Caballé and Daradoumis, 2005d). As a result of this matching, SOA appears to be the best choice to support the development of the most pervasive and challenging collaborative learning environments. In the CSCL context, SOA enhances educational organizations by increasing the flexibility of their pedagogical strategies, which can be continuously adapted, adjusted, and personalized to each specific target learning group. Moreover, SOA facilitates the reutilization of successful collaborative learning experiences and makes it possible for the collaborative learning participants to easily adapt and integrate their current best practices and existing well-known learning tools into new learning goals.

Therefore, in order to increase flexibility and interoperability, the second PSM of the CLPL relies on SOA as it represents an ideal context to support and take advantage of both the latest trends of software development and the benefits provided by distributed systems for the demanding requirements of the CSCL applications to be completely satisfied (see Chapter 4). Using SOA in the context of the CLPL offers the fol-lowing key advantages (Caballé et al., 2007g):

- Simplifies the encapsulation mechanism that is necessary behind a common inter-face of diverse implementations
- Adapts CSCL applications to changing technologies.
- Easily integrates CSCL applications with legacy learning systems and tools.
- Updates pedagogical models and learning tools without causing repercussions on the underlying learning systems and platforms.
- Quickly and easily create and update a learning process from existing services.

Web-services were the implementation technology chosen for this CLPL's PSM given the widely adopted protocols and standards, which represents the rationale of the Web-services approach. These standards represent a suitable context to guarantee interoperability and scalability by taking great advantage of the distributed technologies. This results in a collection of WSDL files organized in directories that are automatically turned into generic, functional Web-services implemented in the desired programming language and allowing developers to implement these services according to specific needs (Caballé et al., 2007e; Caballé, 2008d; See also Figures 3.18 and 3.19). Both the WSDL files and the Web-services of this entire CLPL's PSM are found at http://clpl.uoc.edu/src/clpl-wsdl.zip.

```
<wsdl:message name="groupCreationRequest">
   <wsdl:part name="userSessionId" type="soapenc:string"/>
   <wsdl:part name="csclUserGroupId" type="soapenc:string"/>
   <wsdl:part name="groupFunctioningMark" type="soapenc:string"/>
   <wsdl:part name="groupOutcomes" type="soapenc:string"/>
   <wsdl:part name="groupType" type="soapenc:string"/>
   <wsdl:part name="csclUserCoordinatorId" type="soapenc:string"/>
   <wsdl:part name="csclUserTutorId" type="soapenc:string"/>
   <wsdl:part name="csclUserGroupMembersIds" type="impl:ArrayOf_soapenc_string"/>
   <wsdl:part name="csclWorkspaceId" type="soapenc:string"/>
   <wsdl:part name="groupExtraData" type="impl:ArrayOf_xsd_anyType"/>
</wsdl:message>
```

Figure 3.18: Excerpt of a WSDL file as an example of the service-oriented PSM.

To sum up, the combination of MDA, SOA, and Web-services results in a platform-specific model (PSM) as a collection of WSDL files organized in directories. They are automatically turned into generic web-services by Apache Axis[4], allowing developers to implement the services according to specific needs and using the most appropriate language.

[4]Apache Axis forms part of the Apache Project, found at http://apache.org/axis (Web page as of April 2008)

```java
/**
 * CSCLUserManagement_PortType.java
 *
 * This file was auto-generated from WSDL
 * by the Apache Axis 1.4 Apr 22, 2006 (06:55:48 PDT) WSDL2Java emitter.
 */

package clplws.users.user.business;

public interface CSCLUserManagement_PortType extends java.rmi.Remote {
    public void startUserInterface() throws java.rmi.RemoteException, clplws.users.user.business.CSCLUserManagementException;
    public void closeUserInterface() throws java.rmi.RemoteException, clplws.users.user.business.CSCLUserManagementException;
    public void startUserDiskManager() throws java.rmi.RemoteException, clplws.users.user.business.CSCLUserManagementException;
    public void closeUserDiskManager() throws java.rmi.RemoteException, clplws.users.user.business.CSCLUserManagementException;
    public java.lang.Object groupCreation(java.lang.String userSessionId, java.lang.String csclUserGroupId, java.lang.String groupFunctioningMark, java.lang.Str
    public java.lang.Object groupModification(java.lang.String userSessionId, java.lang.String csclUserGroupId, java.lang.String groupFunctioningMark, java.lang
    public java.lang.Object[] groupsConsultation(java.lang.String userSessionId, java.lang.String[] csclUserGroupIds, java.lang.Object[] groupExtraData) throws
    public java.lang.Object groupsElimination(java.lang.String userSessionId, java.lang.String[] csclUserGroupIds, java.lang.Object[] groupExtraData) throws ja
    public java.lang.Object userCreation(java.lang.String userSessionId, java.lang.String csclUserId, boolean userIsBlock, java.lang.String[] csclUserGroupMembe
    public java.lang.Object[] usersConsultation(java.lang.String userSessionId, java.lang.String[] csclUserIds, java.lang.Object[] userExtraData) throws java.rm
    public java.lang.Object userModification(java.lang.String userSessionId, java.lang.String csclUserId, boolean userIsBlock, java.lang.String[] csclUserGroup
    public java.lang.Object usersElimination(java.lang.String userSessionId, java.lang.String[] csclUserIds, java.lang.Object[] userExtraData) throws java.rmi.
    public java.lang.Object userBlock(java.lang.String userSessionId, java.lang.String csclUserId, boolean userIsBlock, java.lang.Object[] userExtraData) throws
    public java.lang.Object[] treatUserData(java.lang.String userSessionId, java.lang.Object[] userData) throws java.rmi.RemoteException, clplws.users.user.bus
    public java.lang.Object[] treatGroupData(java.lang.String userSessionId, java.lang.Object[] groupData) throws java.rmi.RemoteException, clplws.users.user.bu
}
```

Figure 3.19: An example of a Web-service generated in a specific programming language (Java).

The ultimate aim of the CLPL is to enable a complete and effective reutilization of its generic services and components as the skeleton for the construction of any collaborative learning application, and in particular CSCL applications. Thus, this platform implements the conceptualization of the fundamental needs existing in any collaborative learning experience. In addition, the CLPL is highly interoperable in distributed environments permitting complete flexibility of the services offered in terms of implementation languages and underlying software and hardware platforms.

Chapter 4

Distributed and Grid technology for the enhancement of CSCL

Over the last decade, educational organizations' needs have been changing in accordance with ever more complex pedagogical models as well as with technological evolution resulting in e-Learning environments with very dynamic and changing teaching and learning requirements (Pankatrius, V., Vossen, G. (2003)). In particular, these needs involve extending and moving to highly customized learning and teaching forms in timely fashion, each incorporating its own pedagogical approach, each targeting a specific learning goal, and each incorporating its specific resources. Organizations' demands also include a cost-effective integration of legacy and separated learning systems, from different institutions, departments and courses, which are implemented in different languages, supported by heterogeneous platforms and distributed everywhere, to name some of them (Zaheer et al., 2005).

In addition, collaborative learning environments must provide advanced enablement for distribution both of collaborative activities and of the necessary functionalities and learning resources to all participants, regardless the location of both participants and resources. The aim is to enable the collaborative learning experience in open, dynamic, large-scale and heterogeneous environments (Caballé et al., 2007g).

From this view, one of the main challenges in the development of modern e-Learning systems and in particular collaborative learning applications is to overcome important non-functional requirements arisen in distributed environments such as performance, scalability, flexibility, availability, interoperability, and integration of different, heterogeneous, and legacy learning systems. Specific requirements include:

- wide geographical distribution of learners and tutors who can potentially belong to many different educational institutions,
- access from anywhere, on any learners' computer platform and any software,
- support for a growing load of learning resources and users who access these resources,

- transparent access and share of a huge variety of both software and hardware learning resources,
- multiple administrations from different departments and organizations with specific educational policies,
- inherent dynamism of learners' and tutors' needs and changing learning resources,
- flexibility to reuse pieces of learning resources of different granularity according to specific needs (e.g. customize a curricula by reusing specific learning activities with different input data),
- ease to personalize, update, and meet learning resources by instructors and learners without technology skills.

Finally, as a consequence of the complex knowledge provided to participants in CSCL practices (e.g., constant and automatic learner's assessment according to quantitative and qualitative parameters of the interaction), a huge amount of information is generated from the group activity and gathered usually in data log files (Xhafa et al., 2004). Therefore, there is a strong need for powerful solutions that record the large volume of interaction data and can be used to perform an efficient interaction analysis and knowledge extraction. Grid technology is also increasingly being used in this context to reduce the overall, censored time in processing data by taking advantage of its large computing support (Caballé et al., 2005; Paniagua et al., 2005).

In order to deal with these issues, distributed technology, such as Grid (Foster and Kesselman, 1998) is increasingly used for complex areas, which are computationally intensive and manage large data sets. These features form an ideal context for supporting and meeting the described demanding requirements of collaborative learning applications and, as a result, providing them with important additional benefits, such as wide geographical distribution of resources, multiple administrations from different organizations, transparent access to the resources, and so on.

In this chapter, distributed and Grid technology are taken one step further and presented as a three-fold approach:

1. The combination of Web and Grid services as a basis to help develop Collaborative Learning Management Systems, which may considerable enhance and improve the collaborative learning experience in highly distributed environments (Caballé, 2007b).
2. The use of distributed technology to meet certain requirements which are especially frustrating when they are not fulfilled appropriately during the collaborative learning activity, such as fault-tolerance, scalability, performance, and interoperability (Caballé et al., 2007g).
3. The provision of a Grid-approach for the specific purpose of data intensive computational needs in processing group activity log files and thus allow group learners and tutors to receive relevant and selected knowledge about the group activity even in real time (Caballé et al., 2008a).

Next sections develop each of these three approaches in detail.

4.1 Collaborative Learning Management Systems

In this section, it is proposed an innovative extension of Learning Management Systems (LMS) to the groupware domain, resulting in Collaborative Learning Management Systems (CLMS) (Caballé, 2007b) with the aim of guiding developers in meeting the demanding requirements found in this domain. To this end, the merge of service-oriented infrastructures (SOA), learning management systems, and grids for the support of CSCL. The SOA paradigm (see sub section 1.2.2 for an overview of SOA) plays the key role in this scenario by visioning the entire CSCL needs in terms of flexible, independent, autonomous, interoperable services. To this end, a layered service-oriented architecture for CLMS is proposed containing the main service needs identified in CSCL. This serves as a first step for the development of real CLMS applications by using widely adopted standards, such as those provided by Web-services (see sub section 1.2.2 for an overview of Web-services), which may also benefit the learning process with flexibility and simplicity in the personalization of the pedagogical strategies along with lower costs of integration. Through the exploitation of CLMS in real contexts, it is expected to greatly enhance and improve the collaborative learning experience of all participants of the collaboration.

4.1.1 Web and Grid services

Although SOA can be realized with other technologies, over the last few years Web services has come to play a major role in SOA due to lower costs of integration along with flexibility and simplification of configuration. According to W3C Working Group (2004), a Web service is a software system identified by a URI, whose public interfaces are defined and described using XML. Other systems may interact with the Web service in a manner prescribed by its definition, using XML-based messages conveyed by Internet protocols.

The core structure of Web services is formed by a set of widely adopted protocols and standards W3C Working Group (2004), such as XML, SOAP, WSDL, and UDDI (see Chapter 1 for an overview), which provide a suitable technology to implement the key requirements of SOA. This is so because these protocols allow a service to be platform - and language - independent, dynamically located and invoked, interoperable over different organization networks, and supported by large organizations (e.g. W3C consortium).

Furthermore, when servicing many requests from a highly distributed community, the problem of orchestrating and managing numerous distributed hardware and software components arises. For this reason, the term service-oriented infrastructure is introduced to denote the resource management and provisioning mechanisms used to meet quality of service goals for components and applications (GuiLing, 2005). Grid services come to serve this purpose.

Grid services are essentially Web services with specific extensions or interfaces for use in Grids. Grid services play the central role of the Open Grid Services Infrastructure (OGSI), which intends to provide an infrastructure layer for the Open Grid Services Architecture (OGSA) (OGSI, 2003). At the core of OGSI, a Grid service is a Web service that conforms to a set of conventions for such purposes as service lifetime management, inspection, and notification of service state changes. Grid services provide for the controlled management of the distributed and often long-lived state that is commonly required in distributed applications (Czajkowski et al., 2004).

Since the replacement of OGSI in favor of the Web Service Resource Framework (WSRF) (Czajkowski et al., 2004), Grid services have been replaced by stateful Web services (among other features), and the term Grid service becomes obsolete. WSRF addresses the relationship between Web services and stateful resources through a set of conventions expressed through composable Web services specifications such as the WS-Addressing standard (Czajkowski et al., 2004). However, from the literature, the term Grid service is still on in order to distinguish between those services involved in the core grid level and stateless, typical Web-services implementing grid applications, at the user level. For the remainder of this paper, this view of Grid services will be taken.

4.1.2 Service needs in collaborative learning

This section explores the most common existing e-Learning needs identified in educational organizations. To this end, an overview is provided with the core services required to support collaborative learning applications. The services presented have been chosen by, first, intersecting the most successful e-Learning frameworks and systems such as ELF[1], IAF[2], and OKI [3]. Then, services no specific for CSCL have been omitted. Finally, CSCL-specific services of each framework have been added even though they keep outside the intersection (Caballé, 2007b).

In order to provide a readable, useful set of services, they are grouped together following similar criteria as the frameworks themselves do. Therefore, this section focuses on two main service layers, namely common and application services. In addition, core grid services for e-Learning grids are proposed based on literature.

4.1.2.1 Common services

The services in this sub-section are general purpose so that they may form the basis to any e-Learning environment and may be common across multiple application domains. Common services provide lower-level functionality which is not education-specific, but upon which educational-domain services and users depend:

- *Authentication*. Gather required credentials from an agent, vouches for their authenticity and introduce the agent to the system.
- *Authorization*. Allow an application to establish and query a user's privileges to view, create, or modify application data, or use application functionality.
- *Messaging*. Allow broadcast of messages to users and groups using appropriate communication technology, without being required to understand in advance the specific delivery mechanism that the service implementation will use.
- *Logging*. Enable any other service to be tracked and the corresponding information and events throughout the system are logged for diagnostic, performance, user and, group awareness, feedback, and so on.

[1] E-Learning Framework (ELF) is found at: www.elframework.org (Web page as of April 2008)
[2] IMS Abstract Framework (IAF) is found at: www.imsglobal.org/af (Web page as of April 2008)
[3] Open Knowledge Initiative (OKI) is found at: www.okiproject.org (Web page as of April 2008)

- *Metadata Schema Registry*. Enable access to, and the manipulation of, a registry that apart from meta-data schema typically holds configuration data, application profiles, identifiers or other lookup data.
- *Identifier*. They are responsible for producing and making available learning objects identifiers.
- *Archiving*. Support access to remote storage facilities for storage and retrieval of arbitrary static content.
- *Workflow*. Provide a way to manage an interdependent succession of activities each of which has completion constraints.
- *Search*. Enable the discovery of learning materials and other related information delivered from a system.
- *Service directory*. Hold information about entities such as services, other repositories, people and organizations, and provides support to the finding of available services.
- *Agent*. These are an abstraction of an agent, device, etc. that may include basic information such as id, name, type, role, properties and contact information.
- *User Preferences*. Provide machine-readable information about users' personal preferences, and allows user agents, such as portals, to automatically configure themselves for particular end-users and to prevent end-users from having to enter their preferences into multiple user agents.

4.1.2.2 Application services

The services in this category are educational domain dependent and provide the functionality required by agents. Application services may be implemented so that they have some sort of user interface. Their key requirement is to expose their functionality for reuse by any number of agents or other application services, while implementing a standard interface to support this reuse.

- *Sequencing*. Define the data structures and interfaces responsible for describing the set of possible presentation sequences for the collection of content resources.
- *Content Management*. Provide mechanisms for the creation, flexible management and publishing of content.
- *Assessment*. Support the use of automated assessments. The assessment presentation and reporting is managed at the group and individual level.
- *Grading*. Record the grades, comments, attendance, and scores for a student or group.
- *Group*. Handle the creation, deletion, updating and reading of groups.
- *Member*. Handle the creation, deletion, updating and reading of group members.

- *Course Management*. Handle the creation, reading, updating and deleting of units of learning, courses, modules as well as people information, membership of units of learning, etc.
- *Collaboration*. Abstract service supporting specific synchronous and asynchronous collaboration, such as forum, chat, and whiteboard services.
- *Coordination*. Abstract service supporting specific learning group formation and the definition and planning of the group objectives, such as calendaring and scheduling services.
- *Communication*. Abstract service supporting specific interaction between users, such as e-mail.
- *Awareness*. Abstract service reporting users/groups of what is happening in the learning activity, such as alert, and presence services.

4.1.2.3 Core Grid infrastructure services

From the literature (Sotomayor and Childers, 2006; Ali et al., 2004, and others), there exist many service-oriented infrastructures for use in grids. They provide a cross-domain set of Grid services. By intersecting them with the CSCL and e-Learning domains (see the above-mentioned e-Learning frameworks : ELF, IMS, OKI, and Pankatrius and Vossen, 2003), core Grid services for collaborative learning can be drawn:

- *Sign-on*. Provide authentication, authorization, and access control for data and computing resources.
- *Data Catalog*. Allow datasets to be looked up based on meta-data.
- *Lookup*. Represent the main entry point to access the Grid. They allow dynamic lookup of the Grid services, eliminating the need to know the service locations in advance.
- *Policy*. Set the access rights, rules, and permissions to allow users, agents, and applications to use Grid services.
- *Scheduler*. Resolve a job execution plan for Grid applications. They also submit the plan to the grid for execution.
- *Replica Management*. Add to Grid robustness and scalability by providing the capability to copy and move data around the Grid.
- *Replica Selection*. Locate the optimum replica to use for processing.
- *Steering and optimization*. Allow job requests to adapt to the dynamic environment of the Grid.
- *Monitor*. Provide information on the current state of the job plan. They keep track of the current state of the job and the resources on which jobs are submitted.
- *Accounting*. Enable a fair access to resources as specified by policies.

- *Estimators*. Provide feedback to users, agents, and applications about how much resource a particular action might take.
- *Job Execution*. Execute a set of jobs as part of a job plan.
- *Data Collection*. Provide a way for the application to obtain the final result from the execution of the job execution.

4.1.3 Towards service-oriented collaborative LMS (CLMS) grids

Despite the many e-Learning Grids appeared over the last years, a very few (Bote et al., 2007) are entirely focused on collaborative learning in a service-oriented fashion. In this section, the merge of CSCL, service-oriented e-Learning Grids, and the use of Web and Grid services as implementing technologies, lead to an innovative approach of Collaborative Learning Management Systems (CLMS) for use in grids.

The main reason of creating service-oriented CLMS grids is to produce and consume flexible, interoperable, available, reduced-cost services so as to realize the different pedagogical models designed to fulfill the collaborative learning goals. In addition, these services can be shared and reused by the rest of the organization and cross over different CLMS in the educational sector.

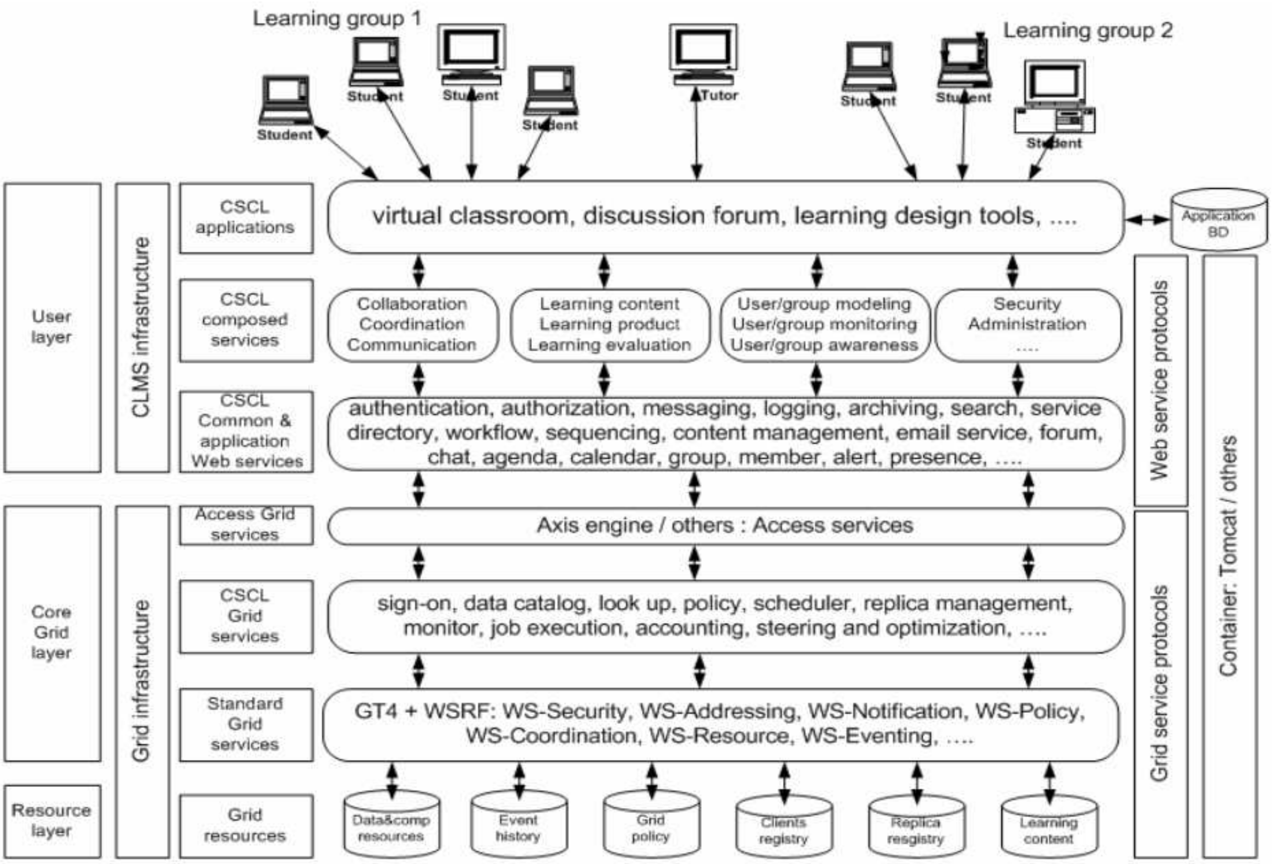

Figure 4.1: Architecture for a CLMS grid.

In Figure 4.1, a fully service-oriented layered architecture for CLMS grids is provided to demonstrate the feasibility of the approach (Caballé, 2007b). This architecture is kept as simple as possible so that technical complexity are hidden (e.g. specific protocols and connectivity issues) and the key aspects can show up. It is made up of two parts: the CLMS and Grid infrastructure. The former is based on the common and application services for CSCL needs described in the last section, which are realized as Web services. The latter consists of both CSCL-specific and standard grid infrastructure services, which are realized as Grid services (i.e. stateful Web-services). Next, the architecture is described in a top-bottom way, using simple terms.

At the upper level, all collaboration actors, equipped with just a Java-enabled browser, see a set of CSCL applications, which they interact with according specific pedagogical goals. Before the collaboration starts, the tutor is in charge of planning and designing appropriate collaborative tasks assisted by specific learning design tools s/he will reach by the browser.

All application functionality is packaged in high level components or abstract services. This will serve to group and organize the whole behavior available as course-grained content-related packages of services so that they can be individually reused and located nearby each other. At the next level, these packaged services are used in the concrete form they were created or with certain degree of composition. In this point, orchestration and choreography standards (W3C Working Group, 2004), such as BPEL and WS-CDL, may enter to play and important role by dynamically form the most suitable grain of service to be used. In any case, an independent service or group of services is to solve a specific users' and system's need, such as authentication, check the agenda, and log the last event.

From this point down, the architecture is about entering the Grid infrastructure. When accessing a grid, some services come to play in the form of external libraries. They are transparent from the application and prepare the current transaction for entering the grid. These services are dependent from the specific environment used to deploy and run the Web services, and are used for administrative, security, and configuration purposes.

In the grid, the appropriate computational and data resources to serve the current CSCL transaction are discovered, replicated, monitored and executed according to a job schedule. All these agents and operations are seen as Grid services (e.g., look up, replica management, monitor, job execution, scheduler, etc.), and thus they are used as services performing a specific CSCL function.

Finally, each CSCL Grid service invokes the suitable service o services provided by the underlying standard framework (i.e. WSRF), which is implemented by the chosen grid middleware platform, such as Globus Toolkit 4 (Sotomayor and Childers, 2006). WSRF provides composable Web services specifications such as the WS-Addressing standard, which provides capability for transport- neutral mechanisms for locating stateful Web services. WSRF is on top of the Resource layer (i.e. Fabric layer in the Grid architecture), and it is responsible for conveniently accessing the resource requested.

From the architectural view, all CSCL support is modeled as services. This provides collaborative learning with high degree of flexibility. No longer have group participants, learning resources, and infrastructure to be tied up in a physical location, but mobility and update are greatly achieved instead.

Standard protocols used by Web and Grid services guarantee maximum interoperability and so legacy and external

CSCL applications can easily join the CLMS and also foster reuse. This allows educational organizations to share their distinct pedagogical models and experiences in the CSCL domain.

On the other hand, the fine-grained service-oriented approach and the use of Web protocols may cause repercussions on the global efficiency of the CLMS. However, the speed-up provided by using grid technology is expected to reduce it and make the system run smoothly.

4.2 Distributed technology for the enhancement and effectiveness of the collaborative learning process

From the experience at the Open University of Catalonia certain requirements are especially frustrating when they are not fulfilled appropriately during the collaborative learning activity, such as fault-tolerance, scalability, performance, and interoperability. They may have considerable repercussions on the learning performance and outcomes as their lack impedes the normal learning flow as well as discriminates learners in terms of technology skills and technical equipment.

- *Fault-tolerance* (Caballé et al., 2007g) refers to provide permanent access to the learning environment and its resources from anywhere and at any time, even in the presence of system failures. The temporary lack of service due to technical difficulties prevents students from fulfilling the learning task in time as well as to meet and collaborate with others as scheduled. Hence, it is expected that the learning system is able to seamlessly recover from failures and keep providing the service.

- A system is told to be *scalable* (Caballé et al., 2007g) if it is able to bear a growing load of both resources and users who may access concurrently to these resources as well as to offer data-intensive and complex functionalities without causing performance repercussions on the underlying system supporting the learning environment. Indeed, a system performing poorly is one of the most frustrating aspects during the on-line collaborative learning experience as it makes participants' requests be waiting for long periods of time to be served. In order to keep on providing a high level of quality of service, a learning system should seamlessly scale to new resources of both hardware and soft-ware at the same pace as the workload increases.

- The gain in *performance* (Caballé et al., 2008e) might help, for instance, include more complex information of the collaboration to be generated and presented in real time (such as modeling the participants' behavior during the discussion by combining individual and group session and navigation information) as well as to make an in-depth analysis through data mining techniques to provide tutors with ongoing progress of students learning during the discussion activity.

- The benefits from having an *interoperable* environment (Caballé, 2007b) is to provide students and tutors with transparent access and share of a large variety of both software and hardware learning resources. Users using any computer platform and software should be able to collaborate among them and interoperate with all the existing resources. Interoperability (Caballé, 2008g) includes supporting users with little technology skills by avoiding the need to set up anything on the client side. Moreover, this may improve the overall learning experience by avoiding the repercussions derived from redundancy and inconsistency of existing databases and information systems in general (e.g., unifying the authentication process so that the user has access to all learning tools by logging in them just once) and integrating seamlessly external and legacy tools and applications.

The aim of these entire improvements is both to enhance the effectiveness of complex collaborative learning processes (e.g., by avoiding a central point of failure) and stimulate the learning experience by describing and predicting students' actions and intentions as well as adapting the learning process to students' features, habits, interests, preferences, and so on. We plan to explore these interesting possibilities in the next iterations of the DF design.

In order to deal with these issues, software techniques and paradigms have been evolving all the time to mainly provide higher levels of abstraction and transparency so that developers can reuse and integrate not only functionality and components but more complex yet larger pieces of software. Indeed, transparency plays a key role in the design of appropriate learning systems that support strong fault-tolerance, dynamic scalability and inherent interoperability, among others requirements. However, although transparency has been greatly enhanced in the e-Learning domain, the barrier of technology incompatibilities and a poor quality of service make the learners' experience ineffective and difficult when interacting with the learning system (Caballé, 2007b).

To that end, on the one hand, service-oriented architectures have come to play a major role in the context of e-Learning due to the benefits that provide in terms of interoperability among heterogeneous hardware and software platforms, integration of new and legacy systems, flexibility in updating software, and so on. On the other hand, distributed technology, such as Grid (see chapter 1 for an overview of Grid) has emerged to extend to a large-scale, flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources (Foster et al., 2001). These features form an ideal context for supporting and meeting the mentioned demanding requirements of collaborative learning applications.

In this section, these entire approaches are taken further through the intensive use of the Collaborative Learning Purpose Library (CLPL) (see Chapter 3 for a complete description) based on fine-grained Web-services (OMG, 2004), especially designed to take great advantage of distributed technology and help develop enhanced collaborative learning systems. The ultimate goal is to provide support for meeting the demanding requirements found in the CSCL domain and considerably improve the effectiveness of the collaborative learning experience.

4.2.1 The CLPL on a distributed infrastructure

In order to fulfill the functionalities designed in the CLPL, the primary principle was to provide a broad set of independent fine-grained Web-services grouped by a particular purpose, such as the authentication process and the presentation of the feedback extracted. The goal was both to enhance the flexibility in the development of CSCL applications and to ease the deployment of these applications in a distributed environment.

To that end, each particular behavior of the CLPL is discomposed into three specialized web-services matching each of the three layers of a typical software development, namely user interface, business and data (Gomaa, 2004; Caballé et al., 2007e). As a result, the completeness of each specific behavior goes through three separate, necessary, sequential steps that connect to the client on one side and to the persistent storage (e.g., database) on the other side.

For instance, the authentication process is formed by three different, independent web-services, namely the authentication user interface, the authentication business, and the authentication data. Thus, when the user tries to log in, the client code calls the authentication user interface web-service, which is in charge of collecting the credentials presented by the user. Then, this web-service calls the authentication business web-service so as to verify the correctness of the user's input (e.g., input no blank, well-formatted, etc.). Moreover, as part of the business process, this web-service validates the users' input upon the information existing in the database by calling the authentication data web-service, which is responsible for accede the data-base and extracting the authentication data of the user.

A clear, independent, and separated vision of each single behavior of the CLPL into fine-grained task-specific web-services results in a natural distribution of the application into different nodes in a network. This distribution is driven by matching the web-service purposes and the node configuration and location in the network. According to this view, the web services in the user interface layer should be allocated nearby the client; the business web-services would be better suited if allocated in those nodes with high-performance processors, and, finally, the data web-services could be attached or nearby the database supported by nodes with high storage capability. As to the database, it can be also distributed as it is clearly separated from the data web-services, which would be in charge of updating and keeping the consistency of the different instances of the database.

The work methodology proposed by the CLPL offers total flexibility as to where (i.e.., network node) to install both each learning system function (i.e., CSCL behavior) and each layer of this function (i.e., web-service). Moreover, the widely adopted standards of the Web-services technology (e.g., HTTP and TCP/IP) help communicate the web-services with each other in a network just using their IP address and passing through firewalls and other barriers that other technologies have problems to over-come. On the other hand, there exist many open-source technologies that deal with Web-services, such as Apache Tomcat and Axis, allowing developers to easily use and deploy the services provided by the CLPL.

In this context, both the independence between the fine-grained services provided by the CLPL and the use of key techniques found in the typical distributed development, such as replication, produce many important benefits. Indeed, by installing and deploying replicas of the web-services all over the network fault-tolerance is easily achieved by redirecting a request to an exact replica of the web-service when a node is down. Concurrency and scalability become natural in this context by parallelizing the users' requests using as many replicas as necessary. Finally, interoperability is inherent in the context of web-services technology as they are fully independent from hardware platforms and programming languages.

To sum up, combining the generic view of CSCL domain provided by the CLPL, the Web-services technology, and leveraging distributed infrastructure, the realization of the most demanding requirements existing in the CSCL environments becomes a reality. In order to validate this approach, Chapter 5 presents a proof-of-concept that takes place in the real context of learning the UOC.

4.3 Efficient embedding of information and knowledge about group activity into CSCL

This section investigates how to efficiently provide effective knowledge extracted from the information collected in Computer-Supported Collaborative Learning (CSCL) environments as essential aspect for any form of cooperation, namely coordination, communication and collaboration (Dillenbourg, 1999). It allows for implicit coordination of collaborative learning, opportunities for informal, spontaneous communication, and gives users awareness (Gutwin et al., 1995) and feedback (Zumbach et al., 2003) about what is happening during collaboration.

Indeed, it is crucial for group members to be aware of others' participation in the collaborative process as this may enhance the collaboration a great deal in terms of decision-making, group organization, social engagement, support, monitoring and so on (Dillenbourg, 1999; Daradoumis et al., 2006). Moreover, providing appropriate feedback about the collaborative activities may impact positively on the motivation, emotional state, and groups' well-being in on-line collaborative learning by means of a steady tracking of parameters related to group functioning, task performance and scaffolding (Daradoumis et al., 2006) and by giving a constant feedback of these parameters to the group. Note that in this context information refers to quantitative and qualitative data generated by the learning group whereas knowledge refers to the result of the treatment of this information in terms of analysis techniques and interpretations that will be presented to the same group that generated it.

Therefore, participants in a collaborative learning experience may greatly enhance their abilities by increasing their knowledge about others in terms of cognitive processes and skills of the students and the group as a whole in solving problems, individual and group effectiveness regarding participation and interaction behavior, social support and help and so on. As a result, the success of CSCL applications depends to a great extent on the capability of such applications to embed information and knowledge of group activity and use it to achieve a more effective group monitoring as well as constantly provide group members with as much awareness and feedback as possible. Awareness (Caballé et al., 2005a) refers to the knowledge provided to participants about both what other participants are doing at the same time and what they did in the past, whereas feedback (Caballé et al., 2005b) goes one step further than awareness by providing exhaustive and elaborated information and knowledge of what is going on in the group over a long period of time. Furthermore, the persistent storage of the knowledge extracted as group memory (Conklin, 1992) is essential for both students and tutors since, on the one hand, it allows participants not to access only the latest documents and data, which are commonly stored for later retrieval, but also the context in which they were created, and, on the other hand, it allows tutors to track the collaborative learning process for several purposes such as scaffolding and assessment of the learning outcome.

In all cases, the provision of effective knowledge implies receiving knowledge simultaneously both synchronously and asynchronously since the current and history interaction data shown are continuously updated. Therefore, on the one hand, users should be aware of the current activity in the group (the contribution of other members, their location and availability, the users working on a shared document at the same time and so on) and should know what other co-participants are doing in real time (e.g. during a multi-user editor session, who is editing and what is being shown). In an asynchronous context, on the other hand, users must know the activities performed by receiving deferred information of who, when, how and where others' interactions have been performed, and also why these interactions have been performed, which implies receiving complex knowledge of the interaction history. However, the supply of efficient and transparent feedback to users in both synchronous and asynchronous modes is a significant challenge. Users are continuously interacting with the system (creating documents, reading others' contributions, etc.) thus generating a lot of events, which, once collected, they must be classified, processed, structured and analyzed (Caballé et al., 2005a; Caballé et al., 2007f). As a consequence of the complex knowledge provided to participants (e.g., constant and automatic learner's assessment according to quantitative and qualitative parameters of the interaction) there is a need for capturing all and each type of possible data that could result in a huge amount of information that is generated and gathered in data log files.

CSCL applications are characterized by a high degree of user-user and user-system interaction and hence generate a huge amount of information usually maintained in the form of event type information. In order to capture the interaction correctly, this event information can be classified into different categories such as work sessions, messages, workspaces, documents and many other objects and thus may generate large size of information, especially, in real online collaborative learning that comprise complex learning activities to be carried out during a rather long period of time and involve a considerable number of participants. This information may include a great variety of types and formats and hence tends to be large in size (Xhafa et al., 2004). Indeed, at a first level of classification, group activity log files of CSCL applications are found associated with synchronous (e.g. multi-user editors) and asynchronous collaboration (e.g. discussion forums). These applications generate different types of information depending on their specific needs and functions (e.g. a discussion forum can generate event-type information so as to capture the participants' contributions). This information can be then stored in different formats.

The experience at the UOC has shown the need to monitor and evaluate real, long-term, complex, collaborative problem-solving situations through data-intensive applications that provide efficient data access, management and analysis. As a result, there is a strong need for powerful tools that record the large volume of interaction data and can be used to perform an efficient interaction analysis and knowledge extraction. Given the real needs of any online collaborative learning situation, in order to provide different types of awareness and feedback, there is a need to capture all and each type of possible data that could result to a huge amount of information that is generated and gathered in data log files. Moreover, the need to make the analyzed information available in real time entails to come across with processing requirements beyond those of a single computer.

As a matter of fact, most of the existing approaches in the literature consider a sequential approach mainly due to three reasons: (i) processing for a specific purpose (i.e. limiting the quantity of information needed for that purpose); (ii) processing the information afterwards (i.e. not in real time) and (iii) processing of small data samples, usually for research and testing purposes (i.e. not for real learning needs). The lack of sufficient computational resources is the main obstacle to processing data log files in real time and in real situations this processing tends to be done later, which as it takes place after the completion of the learning activity has less impact on it (Xhafa et al., 2004). With the emerging Grid technology such a handicap can be overcome by using its computational power. The rest of this section addresses the *sequential versus parallel* approach, which forms the very motivation of this study (see Caballé et al., 2008a for a complete description). Section 5.5 in Chapter 5 presents the implementation of the parallel approach and the results obtained by using Grid technology.

4.3.1 A Sequential Approach

This sub section deals with the problem of extracting useful information from the event logs generated by both the BSCW system and the UOC's virtual campus.

4.3.1.1 The problem of processing log files of the BSCW

For the case of the BSCW, a log processor application in Java was developed, called *EventExtractor*[4]. See Appendix B for a technical overview of this application. This application runs offline on the same machine as the BSCW server and uses the daily log files generated by the BSCW server as input so as to: (i) identify the event boundaries inside the log file, (ii) map specific information contained in these events about users, objects, sessions, etc. to typed data structures, and (iii) store these data structures in a persistent support. In order to analyze the performance of this sequential application, specific test battery was designed in which both large amounts of event information and well-stratified short samples were used consisting of all the existing daily log files making up the whole group activity generated during an academic term in the computer science subject ”Software Development Techniques” at the Open University of Catalonia. This course involved two classrooms, with a total of 140 students arranged in groups of 5 students and 2 tutors. In addition, other tests involved a few log files with selected file size and event complexity forming a sample of each representative stratum and thus obtaining reliable statistical results using an input data size easy to use.

All test batteries were processed by this application on single-processor machines involving usual configurations. The test batteries were executed several times with different workload in order to have more reliable results in statistical terms involving file size, number of events processed and execution time along with other basic statistics. As an example, the experimental results from the sequential processing of the *EventExtractor* application are summarized in Figure 4.2, where for each event log file it is shown the relative comparison scale for the file size, number of events and the processing time. In a similar way, Figure 4.3 presents the processing results of over one hundred event log files involving file size

[4]The EventExtractor source code, tests and results can be found at: http://clpl.uoc.edu/src/EventExtractor.zip (Web page as of April 2008)

and processing time showing that the processing time is linear on the size of the log file processed.

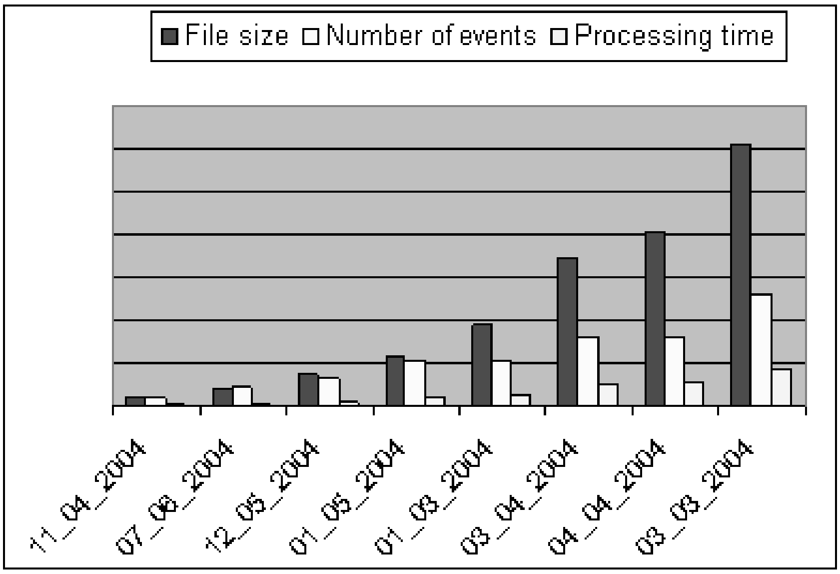

Figure 4.2: Relative comparison scale of a sample of selected log files with the group activity occurring in 8 random days of the spring academic term of 2004. Note that due to the different event complexity, the number of events does not increase linearly with the file size.

However, when executing sequentially in a single machine, this application needs a lot of time and resources to process such amount of information and hence it is not feasible to constantly process data log files in real time. Therefore, these statistical results make evident the lack of computational resources as the main obstacle to constantly present knowledge to users in terms of awareness and feedback in real time.

4.3.1.2 The problem of processing log files of the virtual campus of the UOC

The on-line web-based campus of the UOC is made up of individual and community virtual areas such as mailbox, agenda, classrooms, library, secretary's office, and so on. Students and other users (lecturers, tutors, administrative staff, etc.) continuously browse these areas where they request for services to satisfy their particular needs and interests. For instance, students make strong use of email service so as to communicate with other students and lecturers as part of their learning process.

All users' requests are chiefly processed by a collection of Apache[5] web servers as well as database servers and other secondary applications, all of which provide service to the whole community and thus satisfy a great deal of users' requests. For load balance purposes, all HTTP traffic is smartly distributed among the different Apache web servers available. Each web server stores in a log file all users' requests received in this specific server and the information generated from processing the requests. Once a day (namely, at 01:00 a.m.), all web servers in a daily rotation merge their logs producing a single very large log file containing the whole user interaction with the campus performed in the last 24 hours.

[5]Apache HTTP Server Project is found at: http://httpd.apache.org/ (web page as of April 2008).

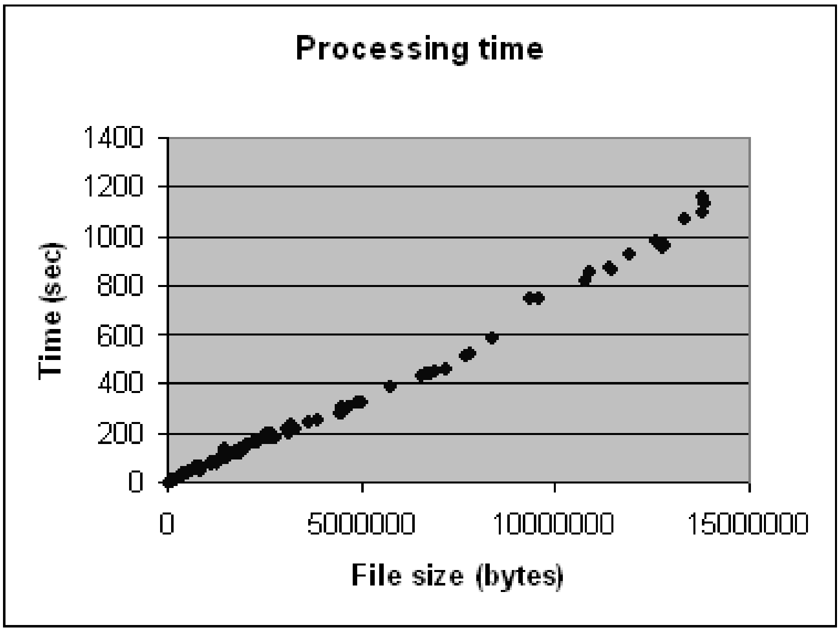

Figure 4.3: Sequential processing time for every event log file size.

A typical daily log file size may be up to 10 GB. This great amount of information is first pre-processed using filtering techniques in order to remove a lot of futile, non relevant information (e.g. information coming from automatic control processes, the uploading of graphical and format elements, etc.). However, after this pre-processing, about 1.8 GB of potentially useful information corresponding to 3,500,000 of log entries in average still remains (Carbó et al., 2005).

Log file entries are structured following a type of format known as Common Log Format (CLF)[6] which is produced by most of web servers including Apache and is fairly configurable. For the purpose of registering the campus activity, log files entries were set up with the purpose of capturing the following information: who performed a request (i.e. user's IP address along with a session key that uniquely identifies a user session); when the request was processed (i.e. timestamp); what type of service was requested (a URL string format description of the server application providing the service requested along with the input values) and where (i.e. an absolute URL containing the full path to the server application providing the service requested).

At this point, it is explained some problems which arose when dealing with these log files. Each explicit user request generates at least an entry in the log file and after being processed by a web server, other log entries are generated from the response of this user request; certain non-trivial requests (e.g. user login) involve in turn requesting others and hence they may implicitly trigger new log entries; the what and where fields contain very similar information regarding the URL strings that describe the service requested and the parameters with the input values; certain information is found in a very primitive form and is represented as long text strings (e.g. user session key is a long 128-character string). Therefore, there is a high degree of redundancy, tedious and ill-formatted information as well as incomplete as at some cases certain user actions do not generate any log entry (e.g. user may leave the campus by either closing or readdressing the browser)

[6]Common Log Format: http://httpd.apache.org/docs/1.3/logs.html#common (web page as of April 2008).

and have to be inferred. As a consequence, treating this information is very costly in terms of time and space needing a great processing effort.

In order to deal with the above mentioned problems and inconveniences, an *ad hoc* application in Java was developed called *UOCLogsProcessing*[7] that processes log files of the UOC (see Appendix B for a technical overview). In particular, this application runs offline on the same machine as the logging application server. It uses, as an input, the daily log files obtained as a result of merging all web servers' log files. The following process is run: (i) identify the log entries boundaries and extract the fields that make up each entry, (ii) capture the specific information contained in the fields about users, time, sessions, areas, etc., (iii) infer the missing information, (iv) map the information obtained to typed data structures, and (v) store these data structures in a persistent support.

However, after running similar test batteries, the repercussions of processing UOC's log file data sequentially are very similar to those mentioned for the case of BSCW log file processing. Thus, the sequential processing of the *UOCLogsProcessing* also takes too long to complete the work and it has to be done after the completion of the learning activity, which impedes from providing complex feedback to users in real time (Xhafa et al., 2004).

An improved version of the processing of log files is presented next to parallelize the processing of information and the main experimental results achieved. Although this approach is generic and is valid for both types of log files (i.e., the BSCW system and the UOC virtual campus), for testing purposes it is used the type of log file which most fits and can take advantage of characteristics of each resulting prototype.

To this end, next subsections report the different experiments carried out at the UOC that show the feasibility of the Grid approach to achieve an effective embedding of the appropriate knowledge into collaborative learning practices. First it is shown the context that motivated this research and some reference are made to other studies in the field. Next subsection presents the process of creating effective knowledge and exemplify a real situation. Then, it is proposed a general structure of data log file used for gathering the event information generated during group activity. This motivates the introduction of the Master-Worker approach for processing log files using Grid infrastructure. The implementation of this general approach using real Grid infrastructure and the analysis of the results obtained are described in Chapter 5.

4.3.2 Introducing a parallel approach

Several studies have been conducted at the UOC (Caballé et al., 2005b; Paniagua et al., 2005; Caballé et al., 2007a; Caballé et al., 2007c; Caballé et al., 2007d; Caballé et al., 2008a; Caballé et al., 2008e) to show that a Grid approach can increase the efficiency of processing a large amount of information from group activity log files. These studies have involved the interaction data collected from the log files of both the BSCW system (Bentley et al., 1997) used at the UOC to support Problem-Based Learning practices in small groups and the own virtual campus of the UOC. The experimental results shows, first, the gain provided by the Grid approach in terms of relative processing time and, second, the benefits of using the inherent scalable nature of Grid while the input log files are growing up in both number and large size.

[7] The UOCLogsProcessing source code, tests and results can be found at: http://clpl.uoc.edu/src/UOCLogsProcessing.zip (Web page as of April 2008)

The initial real context of this study refers to group activity at the UOC and the Math Forum at Drexel University[8] (USA). The former results from applying the Project-Based Collaborative Learning paradigm to model several online courses, such as ”Software Development Techniques”. These courses involve hundreds of students, a dozen of tutors and are characterized by intensive collaboration activity due to the complexity of the learning practices.

To implement the collaborative learning activities and capture the group interaction the BSCW system is used, a groupware tool that enables asynchronous collaboration over the web. BSCW records the interaction data into log files, which can be used for interaction analysis and knowledge ex-traction. However, its centralized architecture does not allow data access, management and the analysis of BSCW log files. In particular, BSCW does not incorporate functionalities to process the log files nor provides the means to calculate and present statistics results. Moreover, BSCW generates a unique log file at the end of the day, which includes a large volume of data describing the activity of all virtual groups. Given that log files do not classify or structure information in any way, there is no possibility of scaling them up. As a result of this there is no way to access data related to separated workspaces, specific groups, or phases of the learning practice.

In recent years the popularity of distant collaborative learning among students has increased enormously. Typically, in an academic term more than 500 students are distributed into more than 100 virtual groups composed of 4 to 6 members. All the groups work, mainly asynchronously, during 4 months. Due to the large volume of interaction data generated, the wide geographical distribution of the students, and the limitations of BSCW, a Grid solution for data-intensive applications and data analysis becomes imperative to overcome the above-mentioned problems and provide a more effective service to students and tutors.

Similarly, the VMT Project has been investigating how small groups of students meet online and solve mathematical problems collaboratively using Synergeia (Stahl, 2002) (an extension of BSCW) and other similar systems. This study gives great importance to the processing of transaction logs of the collaboration activities, which is currently done manually. As the size of transaction logs increases, it will become even more necessary to develop a means for the automatic processing of data.

In the context of this research, Grid computing (Foster and Kesselman, 1998) has been used to support the real-time requirements imposed by human perceptual capabilities as well as the wide range of many different interactions that can take place as one of the most challenging issues of collaborative computing support.

There is currently a lot of research being conducted on how to use the Grid computing paradigm for complex problem solving (Foster and Kesselmann, 1998), processing huge amount of data in biology and medicine, simulations, and collaborative systems. For such problems, putting together distributed computing and storage resources is clearly of great value. Moreover, different technologies such as Globus[9], MPI-Grid2 —footnoteMPICH-G2: A Grid-enabled MPI.

[8]The Virtual Math Teams (VMT) Project at Drexel University aims to develop the first application of digital libraries to small group collaborative learning, which requires the processing of a large volume of collaborative activity log files. For more information, see http://www.cis.drexel.edu/faculty/gerry/vmt/index.html (web page as of April 2008)

[9]Globus: http://www.globus.org (web page as of April 2008)

http://www3.niu.edu/mpi/ (web page as of April 2008), Condor-G[10], NetSolve (Casanova and Dongarra, 1998) and frameworks such as Master-Worker Framework on computational Grid (Goux et al., 2000) as well as infrastructures for data-intensive Grid applications (Pérez et al., 2003) have been proposed to support the development of Grid-based applications.

For instance, Grid computing offers high-throughput and data-intensive computing (Caballé et al., 2004), which greatly facilitate the process of embedding information and knowledge into CSCL applications making it possible to provide users with real-time awareness and constant feedback. In the literature, however, there has been little study aimed at achieving these objectives. As an initial approach, the OCGSA framework (Amin et al., 2002) pro-poses an event archiving service, which logs the messages or events communicated between online users of a group instance into a persistent database. However, in proposing the implementation of the functionality, this framework does not offer any methodology which takes advantage of the distributed nature of Grid computing to partition the generated event information for efficient parallel processing.

4.3.3 The structure of group activity log files

In collaborative learning systems, usual group activity results in a lot of interaction which generates a huge amount of events. Therefore, CSCL applications have to be designed to permit the pre-structuring, classification and partitioning of these large amounts of event information into multiple log files to meet different criteria (e.g. group or time) in order to correctly capture the group activity and increase the efficiency of data processing.

The existing CSCL applications have several drawbacks in structuring the log files that prevent efficient processing. To overcome this, it is firstly proposed a definition and classification of event information generated in a CSCL system and, secondly, explained how to store this information in log files according to different criteria with the aim of facilitating its later processing in a Grid infrastructure (see Xhafa et al., 2004 and Figure 2.2 in Chapter 2 for this subsection).

4.3.3.1 Definition and classification of event information

The most important issue while monitoring group activity in CSCL applications is the collection and storage of a large amount of event information generated by the high degree of interaction among the group participants. Such a large amount of informational data may need a long time to be processed. Therefore, collaborative learning systems have to be designed in a way that pre-structures and classifies information in order, on the one hand, to correctly measure the group activity and, on the other hand, to increase the efficiency during data processing in terms of analysis techniques and interpretations.

The classification of information in CSCL environments is achieved by distinguishing three generic group activity parameters: task performance (i.e. collaborative learning product), group functioning and scaffolding (Caballé et al., 2004). Furthermore, in a collaborative learning experience, the group activity is driven by the actions of the participants on the collaborative learning resources, which are aggregated to the user events to form another taxonomy distinguishing,

[10]Condor-G: http://www.cs.wisc.edu/condor/condorg/ (web page as of April 2008)

at a high level of abstraction, between active, passive and support user actions (see Chapter 2 and Caballé et al., 2007f). Therefore, in CSCL applications there is a strong need for the classification of all types of events generated by user actions according to the three generic parameters mentioned. To this end, a complete and tight hierarchy of events (see Figure 2.1 in Chapter 2) is provided to collect and categorize the identified events generated by user actions during the collaborative learning activity.

4.3.3.2 The structure of the log files

In order to prepare the event information for efficient processing, as soon as it is classified and turned it into persistent data, this information is stored in the system as log files, which will contain all the information collected in specified fields. Next, it is intended to predefine two generic types of log files according to the two basic criteria, time and workspace, that characterize group collaboration. These log files will represent as great a degree of granularity as possible regarding both criteria and they will be parameterized so that the administrator can set them up in accordance with the specific analysis needs. Thus, the finest grain or the smallest log file should be set up to store all events occur-ring in each group for the shortest time interval. Therefore, every single workspace will have its own log file made up of all the events occurring within the workspace for a given period of time.

During data processing it will be possible to concatenate several log files so as to obtain the appropriate degree of granularity thus making it possible for a distributed system to efficiently parallelize the data processing according to the characteristics of the computational resources. The aim is to efficiently process large amounts of information enabling the constant presentation of real-time awareness and constant feed-back to users during the group activity.

Thus, concatenating several log files and processing them in a parallel way, it would be possible to constantly show each group member’s absolute and relative amount of contribution, which would provide participants with essential feed-back about the contribution of others as a quantitative parameter supporting the production function. In a similar way, real-time awareness is possible by continuously parallelizing and processing each and every single fine-grained log file of each workspace involved at the same time in order to permanently notify all workspace members of what is going on in their groups. Finally, showing the results of complex statistics after longer periods of time (e.g. at 12 hour intervals) is very important for the group’s tutor to be able to monitor and assess the group activity as a qualitative parameter supporting acquisition of information about students’ problem-solving behavior, group processing and performance analysis.

4.3.4 A Master-Worker approach for processing log files

The Master-Worker (MW) model (also known as Master-Slave or Task Farming model) (Goux et al., 2000) has been widely used for developing parallel applications. In the MW model there are two distinct types of processors: master and workers. The master processor performs the control and coordination and assigns tasks to the workers. It also decides what data will be sent to the workers. The workers typically perform most of the computational work. The MW model has proved to be efficient in developing applications using different degrees of granularity of parallelism. Indeed, it has several advantages such as flexibility and scalability (the worker processors can be implemented in many different ways and they can be easily added if needed) as well as separation of concerns (the master performs coordination tasks and the worker processors carry out specific tasks). This paradigm is particularly useful when the definition of the tasks to be completed by the workers can be done easily and the communication load between the master and workers is low. (see Xhafa et al., 2004 and Figure 2.2 in Chapter 2 for this subsection)

4.3.4.1 Master-Worker paradigm on the computational grid

The MW paradigm has been used in developing parallel applications in traditional supercomputing environments such as parallel machines and clusters of machines. Over the last few years, Grid computing has become a real alternative for developing parallel applications that employ its great computational power. However, due to the complexity of the Computational Grid, the difficulty encountered in developing parallel applications is higher than in traditional parallel computing environments. Thus, in order to simplify the development of Grid-aware applications several high-level programming frameworks have been proposed, among which is the Master-Worker Framework (MWF) (Goux et al., 2000).

MWF allows users to easily parallelize scientific computations through the master-worker paradigm on the computational grid. On the one hand, MWF provides a top level interface that helps the programming tasks to distribute large computations in a Grid computing environment; on the other hand, it offers a bottom level interface to existing grid computing toolkits, for instance, using the Condor system to provide Grid services. The target applications of MWF are parallel applications with weak synchronization and reasonably large granularity. This framework shows its suitability for processing log files of group activity since different degrees of granularity available are considered to guarantee efficiency and, furthermore, there is no need for synchronization or communication between the worker processors. Moreover, in the application, the communication load between the master and workers is very low.

4.3.4.2 The architecture of the application

The architecture of the application (see Figure 4.4) is made up of three parts: (1) the Collaborative Learning Application Server, which is in charge of maintaining the log files and storing them in specified locations; (2) the MW application for processing log files and, (3) the application that uses the resulting information in the data bases to compute statistical results and present them to the final user.

The Master-Worker Application for Processing Log Files. Here, the aim is to present more details of the MW application, basically how the master and worker processors are programmed. The master is in charge of generating new tasks and submitting them to the MWDriver for distributing them to the worker processors while the worker processors run in a simple cycle: receiving the message describing the task from the master, processing the task according to a specified routine and sending the result back to the master. The MW framework, which schedules the tasks, manages the lists of workers and of tasks to be performed by the MWDriver. Tasks are assigned to workers by giving the first task on the list to the first idle worker on the worker list. The fact that the MWDriver's interface allows the task list to be ordered

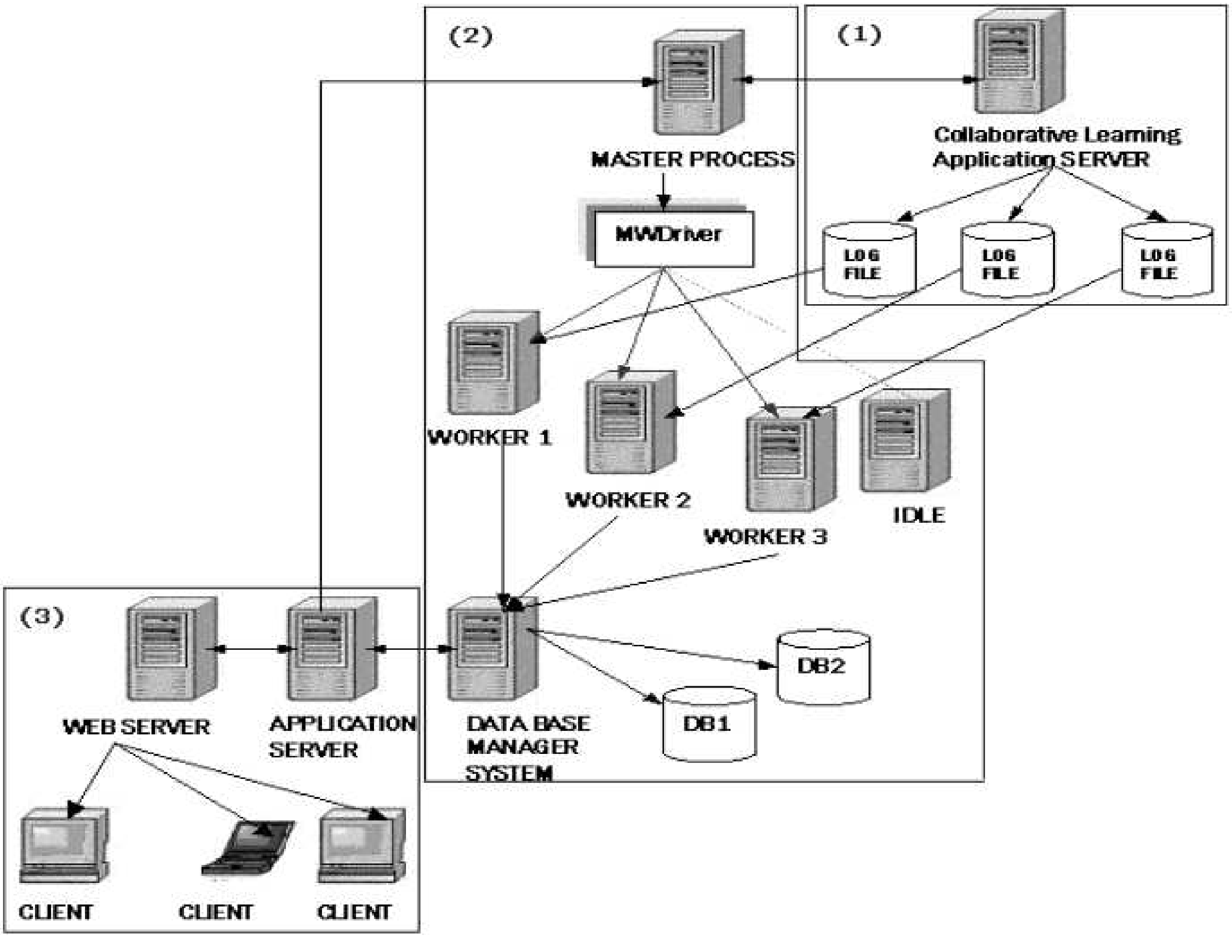

Figure 4.4: The architecture of the application for processing log files.

according to a user's criteria is considered as well as the list of workers to be ordered according to their computational power. Thus, the task list is ordered in decreasing order of log file size and the machines in decreasing order of processing capacity so that "good" machines have priority in receiving the largest log files.

Furthermore, a unique type of task to be performed by the workers consists in processing a log file. It is assumed that the workers have the processing routine available; otherwise, the worker would take a copy of the routine on receiving a task for the first time and then use a flag to indicate whether it must receive a copy of the routine or not. The task is described as follows:

- *Task description*:

```
address of the location of the log file;
name of the log file;
size of the log file;
address of the location where the processing routine is found;
url of the database where the processed information will be stored;
```

- The *master processor* is programmed as follows:

```
while (true) do
check for new log files generated from the Collaborative Learning Application
Server;
update the list of the <log file description> for the new incoming log files;
for each new log file generate a task;
submit the newly generated tasks to the MWDriver;
```

Please note that the log files generated by the Collaborative Learning Application Server can be stored either in disk spaces of the same server or at different locations (machines) available in the Grid. Furthermore, the processed information by the workers can be stored either in unique or different databases that can be found at different machines as specified in the tasks to be realized by the worker processors.

- The *worker processor* is programmed as follows:

```
receive the task;
receive the specified log file from the specified location in the task descriptio
run the processing routine on the log file;
```

```
send to the master the task’s statistics (execution time, number of events proces
upon completion of the task;
```

Efficiency issues of the MW Application. It should be observed that the communication takes place between master and the workers at the beginning and the end of the processing of each task. Therefore, the application has weak synchronization between the master and the workers, which ensures that it can run without loss of performance in a Grid environment. Moreover, the number of workers can be adapted dynamically so that if new resources appear they can be incorporated as new workers in the application; in addition, if a worker in the Grid becomes unavailable while processing a task, the task can be reallocated to another worker. Finally, by having different degrees of granularity of the log files it is possible to efficiently distribute the load balance among workers and minimize the transmission of the data log files from their original locations to the worker machine.

4.3.5 The design of the resulting database

Once the event information from the log files has been processed, the workers (see Figure 4.2) send back the task reports (e.g. processing time, number of processed events, etc.) to the collaborative learning application server through the master so as to verify the results achieved. The results of data processing, which workers send to the data-base manager system, should have correctly represented all the information contained in the log files so as to make it possible to consult both the desired data from the data-base directly (e.g. number of connected users, type of documents in a certain workspace, etc.) and the computed complex statistical results from the database. These statistical results should be obtained by the application server as fast as possible and presented to group members and tutors in different formats.

Thus, based on the premises argued in (Watson, 2003), it is provided a logic design of the database[11], which is generic, efficient and independent from any specific database manager. The database is designed in a way to satisfy all of these requirements and to allow users to consult data regarding the basic entities that take place in any CSCL environment (users, objects, workspaces, connections, etc.).

4.3.6 XML representation of the statistical results

The third part of the application uses the resulting information in the databases to compute statistical results and present them to the members of the online collaborative group and the tutors In this context, current research is studying an XML coding of the statistical results in order to make it possible to present this information to final users in different forms. Considering the fact that the data is highly structured and the design of the relational database (Fernández et al., 2002), it is proposed that application be designed as a middleware (Kyung-Soo, 2001) which performs the following functions: to extract necessary information from the databases, to compute statistical measurements as desired, and to convert the results into XML output. This design will provide sufficient flexibility as to allow ad hoc statistical measurements to

[11] See the design of the database at: http://clpl.uoc.edu/docs/DBDesign.pdf

be obtained as well as permitting the creation of user-specified document type definitions (DTD) to accommodate the different needs of information representation.

Chapter 5

Merging CSCL, generic programming and distributed computing: Evaluation and results

Based on the conceptual model of interaction management in CSCL presented in Chapter 2, the generic CLPL computational model introduced in Chapter 3 and the distributed approach of Chapter 4, several prototypes of an effective structured and context-aware asynchronous discussion forum for collaborative knowledge building were developed and used to support real learning experiences in the context of the Open University of Catalonia (UOC)(Caballé and Xhafa, 2005c; Caballé et al., 2007e; Caballé et al., 2007g; Caballé et al., 2008b; Caballé et al., 2008e; Caballé and Feldman, 2008f; Caballé et al., 2008h).

These novel experiences are reported in these sections from all stages of their development and experimentation, which give new opportunities to learning by discussion, and is applied to meet new pedagogical needs. To this end, a discussion and reasoning process is first described briefly in the form of requirements and then the development of the two applications and their effects in real the learning processed happening at the UOC are reported in detail. Then, distributed infrastructure is added to meet non-functional requirements (Caballé et al., 2007g; Caballé et al., 2008a). Finally, the results of using the CLPL for developing small CSCL tools in an effective and timely fashion are also reported.

5.1 Preliminaries

About 740 graduate and undergraduate students were involved either directly or indirectly in 6 experiences that took place at the UOC last 3 academic terms (i.e., Spring 2007, Fall 2007 and Spring 2008) as part of the curricula of several courses, namely Methodology and Management of Computer Science Projects (MGPI), Management of Organizations and Computer Science Projects (GOPI), and Software Engineering (EP).

The design of both the collaborative tools and the discussion assignments described in this chapter followed the new educational principles and paradigms proposed by the Bologna Process as a basis of the new European Higher Education Area (Kulesza and Reinalda, 2006), which are supported by well-fundamented pedagogical theories (e.g., constructivism (Jonassen, 1994)). These principles consider students as active and central actors in their learning process. On the other hand, lecturers are moving from passive knowledge transmission agents to specialist agents who design the course, guides, assists and supervises the student's learning process (Simonson et al., 2003). From this view, the experiences here reported are fully student-centered leaving the lecturer as a supportive actor who no longer interferes in the collaboration at his/her convenience but provides adequate scaffold to enhance and improve knowledge building as a constructive process among learners.

For space reasons only two out of these six experiences are described in detail in this chapter. Results from these and the rest of experiences are shown in Appendix A. Even though these two experiences are representative from the research and pedagogical standpoint, all of them performed very similarly in terms of dynamics of the discussion and impact on both quantitative and qualitative participation as well as their behavior concerning certain technical issues to deal with. Therefore, to certain degree, the results and their analysis explained in this chapter may be extrapolated to the whole set of experiences.

The entire experiences were preformed with different versions of two major software development efforts, namely the Discussion Forum (Caballé et al., 2008b) and Communities of Learning Practice Environment (Caballé et al., 2008f). See Appendix B for a technical view of these two applications. Both tools took great advantage of the Collaborative Learning Purpose Library (CLPL) (see Chapter 3 and Caballé et al., 2007e for an extensive overview) platform, which was used extensively in all stages of the their development. As a summary, the CLPL is made up of five components in all handling user management, security, administration, knowledge management and functionality and enables a complete and effective reutilization of its generic components for the construction of specific CSCL applications. The aim is both to map the essential elements involved in any CSCL collaborative learning application and support the conceptual model of interaction analysis explained in Chapter 2. To this end, this library is mainly performed by the two components, namely *CSCL Knowledge Management* and *CSCL Functionality components*, which form the core of the computational model in the construction of collaborative learning applications. The *CSCL Knowledge Management* component, in short, manages the system log files made up of all the events occurring in a certain workspace over a given period of time and performs the statistical analysis event information as well as the management and maintenance of the knowledge extracted by that analysis. The *CSCL Functionality* component, which has five subsystems in all, defines the three elemental parts involved in any form of cooperation, namely coordination, communication and collaboration. This component also realizes the presentation to users of the knowledge extracted by the previous component in terms of immediate awareness and constant feedback of what is going on in the system.

Other important development efforts include two Java-based log files processors, namely *EventExtractor* and *UO-*

CLogsProcessing. See Chapter 4 and Appendix B for details of these two applications. The former processes log files of the BSCW while the latter processes and analyzes information from the log files of the virtual campus of the UOC. Parallel versions of the same applications were also developed to take advantage of distributed and Grid infrastructure. These tools represent the very rationale of the Grid approach presented later on in this chapter.

5.2 Discussion Forum: The development of an effective knowledge-based structured discussion tool

Based on the conceptual model of interaction management presented in Chapter 2, a prototype of a web-based knowledge-based structured collaborative learning system, called Discussion Forum (DF), was developed to support and foster collaborative knowledge building by means of the discussion process (Caballé et al., 2008b; Caballé et al., 2008h). To this end, a discussion and reasoning process is first described briefly in the form of requirements and then the design of the application is treated in certain detail.

5.2.1 Pedagogical background and requirements

Requirements[1] in collaborative learning environments involve the discussion process, which forms an important social task where participants can think about the activity being performed, collaborate with each other through the exchange of ideas arising, propose new resolution mechanisms, and justify and refine their own contributions and thus acquire new knowledge (Salomon, 1993).

To this end, a complete discussion and reasoning process is proposed based on three types of generic contributions, namely specification, elaboration and consensus. Specification occurs during the initial stage of the process carried out by the tutor or group coordinator who contributes by defining the group activity and its objectives (i.e. statement of the problem) and the way to structure it in sub-activities. Elaboration refers to the contributions of participants (mostly students) in which a proposal, idea or plan to reach a solution is presented. The other participants can elaborate on this proposal through different types of participation such as questions, comments, explanations and agree/disagree statements. Finally, when a correct proposal of solution is achieved, the consensus contributions take part in its approval (this includes different consensus models such as voting); when a solution is accepted the discussion terminates (Stahl, 2002).

Finally, in a discussion process, participants perform a role according to their profile (e.g. coordinator, member, guest, etc.), have personal collaborative preferences (e.g., language) and must set up environment features (e.g. sound or visual effects, text or voice warnings, etc.) according to their personal characteristics. Participant needs are not static and they evolve as the discussion moves forward (Caballé et al., 2004).

It is worth mentioning here that the decision to develop a newly entire system rather than simply adding a module

[1]See the complete Discussion Forum's requirements and their analysis at: http://clpl.uoc.edu/docs/DiscussionForumSpecification.pdf (Web page as of April 2008)

for the purposes to some existing e-learning system came in order to overcome important non-functional requirements arisen in this context, especially in distributed environments. Indeed, non-functional requirements, such as user scalability, resource availability, performance, interoperability, and integration of different, heterogeneous, and legacy collaborative learning systems, may have considerable repercussions on the collaborative learning performance and outcomes when they are not fulfilled appropriately. Failing to meet these important requirements impedes the normal learning flow as well as discriminates learners in terms of technology skills and technical equipment (Caballé et al., 2008c). For instance, despite having a web-based collaborative learning system with advanced functionalities installed in a server, when the server is down for technical reasons, all participants have to temporarily stop participating, which causes great deal of frustration, especially at the UOC, where students have very limited time to dedicate to their studies. The lack of existing collaborative learning applications featuring these non-functional aspects encouraged to entirely build the innovative DF system.

5.2.2 The design of the application

The design[2] of the DF includes certain thematic annotation tags based on the low-level exchange categories identified in Chapter 2 (see section 2.2), such as information-clarification and request of opinion (see Table 5.1) for a list of all categories), which qualifies each contribution and as a result structure the discussion process. In order to avoid unnecessary choice, each context of the discussion process determines a precise and short list of just those categories that are possible in a certain point of the discussion process (e.g., in replying any kind of request, just the cards involving the provision of information are provided to classify the reply). This makes the choice of the appropriate tag much shorter and easier and no error-prone (see Figure 5.1). In addition, as part of the design, the tutor is to examine and assess all contributions based partially on the tags used by students to categorize them, and as a result students are aware of the potential repercussions of tagging posts incorrectly in order to optimize the assessment instead of reflecting the true meaning of their posts.

Consequently, all contributions are recorded as exchange acts, analyzed and presented as information to participants either in real time (to guide directly students during the learning activity) or after the task is over (in order to understand the collaborative process). To this end, the CLPL's *CSCL Knowledge Management* component provided full support to the interaction management. In particular, a complete treatment of the structured interaction generated enabled the system to keep participants aware of the contributing behavior and dynamics of others, to check certain argumentative structures during discussion and assist in achieving a more satisfactory solution to the problem during the consensus phase, and finally to provide feedback based on the data produced.

CLPL's *CSCL Functionality* component provided suitable support in the design of the virtual places where the discussions take place. Indeed, the *room* entity was recursively used in different levels of abstractions, such as folders to hold the assignments featuring the class discussions and discussion threads inside each discussion. This also eased the implementation by reusing the same code for both purposes. This component also provided the suitable means to present the

[2]The complete design of the Discussion Forum is found at: http://clpl.uoc.edu/docs/DiscussionForumDesign.pdf (Web page as of April 2008)

Table 5.1: List of categories to classify a contribution.

Moves	**Categories**
support	Greeting
	Encouragement
	Motivation
request	REQUEST-Information
	REQUEST-Elaboration
	REQUEST-Clarification
	REQUEST-Justification
	REQUEST-Opinion
	REQUEST-Illustration
inform	INFORM-Extend
	INFORM-Lead
	INFORM-Suggest
	INFORM-Elaboration
	INFORM-Explain/Clarification
	INFORM-Justify
	INFORM-State
	INFORM-Agree
	INFORM-Disagree
set-up-an-issue	PROBLEM-Statement
provide-solution	PROBLEM-Solution
consent-solution	PROBLEM-Extend solution
	PROBLEM-Assent solution

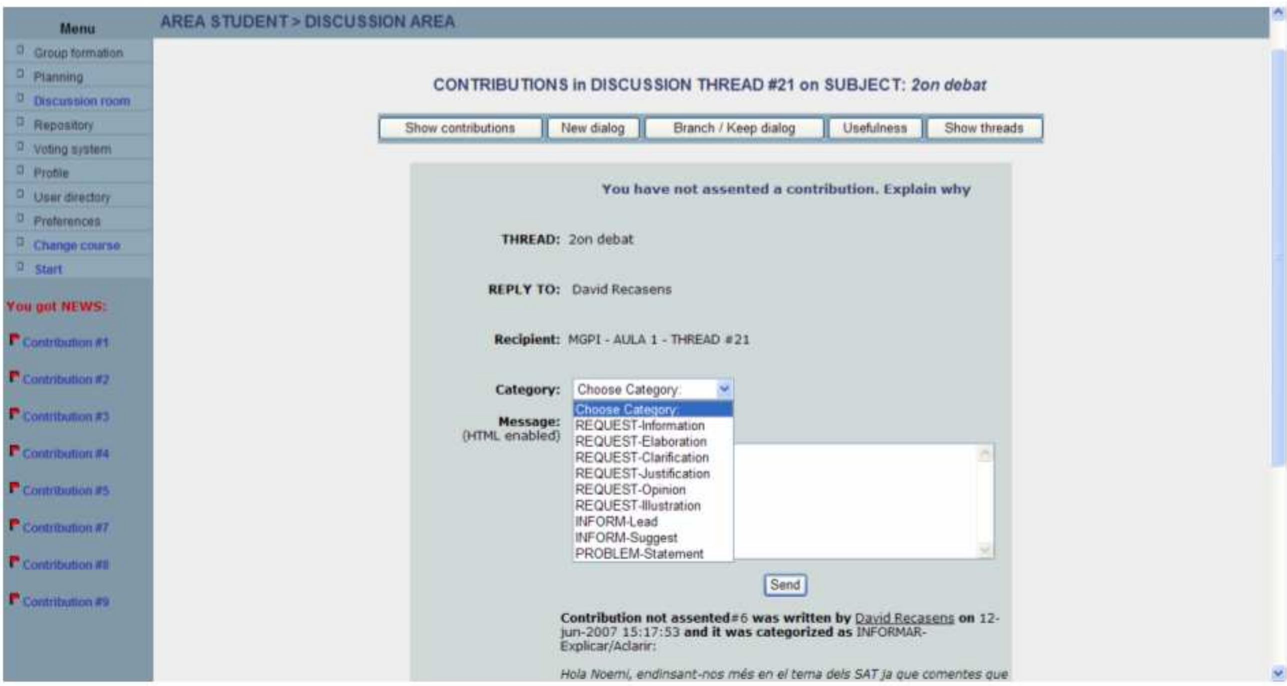

Figure 5.1: The specific list of cards for a reply to a contribution categorized as INFORM-Explain.

information of the knowledge acquired from the data interaction to the participants in the form of appropriate awareness and feedback, representing the cornerstone of this approach. The ultimate aim is to achieve a more effective interaction by allowing all participants to be aware of both their own and others' performance during the discussion process. See the DF development at to acquire a complete knowledge of the level of reusability achieved from this component.

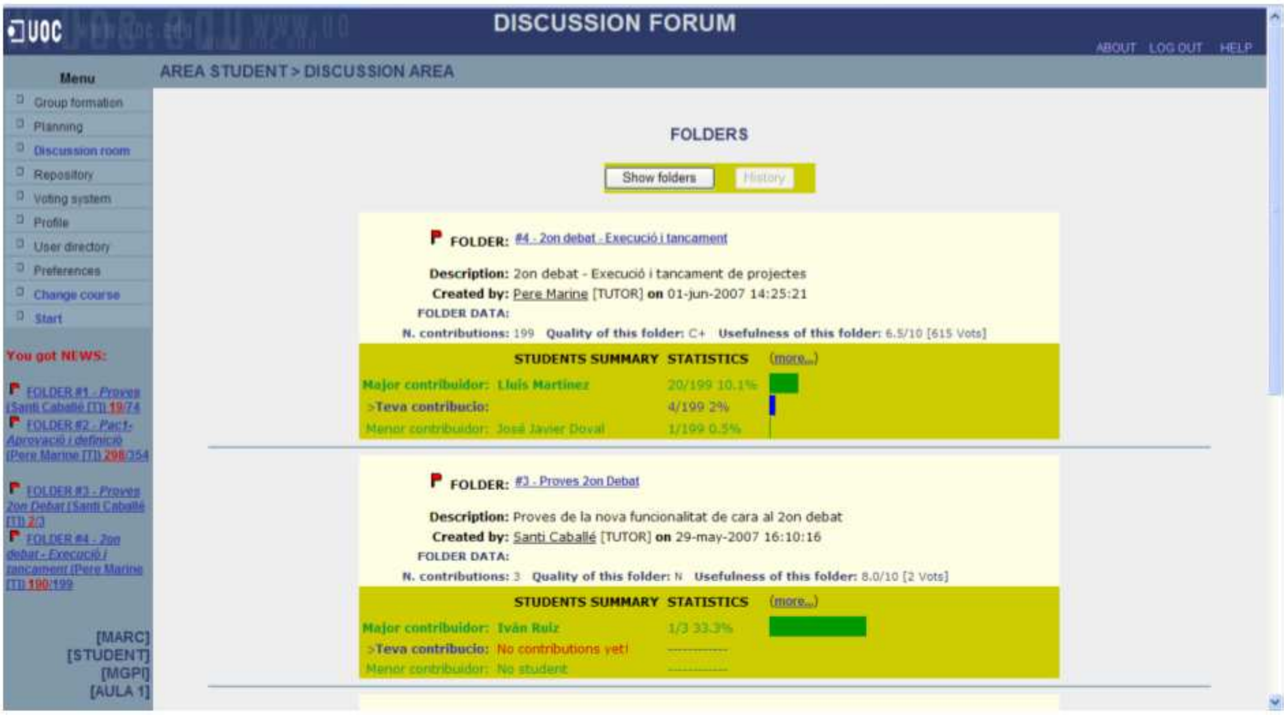

Figure 5.2: A snapshot of awareness and simple quantitative feedback provided.

Therefore, the DF was especially designed to provide students with additional and important features to support the discussion in comparison to the traditional, well-known discussion tool used in the virtual classrooms of the UOC , such as (i) updated feedback, which includes the current mean number of all contributions' (see Figure 5.2) and complex indicators about the collaboration (see Figure 5.3), (ii) threads in fully separated rooms (see Figure 5.4), (iii) open-closed branched dialogs (see Figure 5.5 and 5.6), and (iv) contribution qualifiers (see Figure 5.1). Consequently, DF's users were urged to qualify their contributions (using the annotation cards of Table 5.1 before sending a new or reply post as well as to decide whether their contribution closed the current dialog.

In particular, participants' contributions in each thread were designed as structured dialogs with the aim of separating the different types of low-level exchanges. Moreover, dialogs are to be closed when a request is satisfied or a basic problem is solved (see Figure 6a) and to be branched from a specific exchange (i.e. problem-statement) so as to provide different solutions to the same statement (see 5.6). Finally, a contribution is both to be assented depending on the context and to be evaluated by the other participants in terms of utility in their progress in the discussion (see Figure 5.7).

A discussion process is conducted by a tutor who continuously monitors the discussion threads with the aim of both assessing the contributions and providing support when needed by posting clarifying contributions in any thread and/or starting supporting threads. The contribution assessment by the tutors is performed in a very similar way as the peer

Figure 5.3: A snapshot of complex and updated feedback provided to all participants. In this case, student Marc is located in the 19th position in the rank.

Figure 5.4: A snapshot of some discussion threads inside a folder holding the discussion.

Figure 5.5: An open-closed complete dialog inside a discussion thread.

Figure 5.6: Two dialogs in the same thread; upper dialog can be branched by checking the first check box.

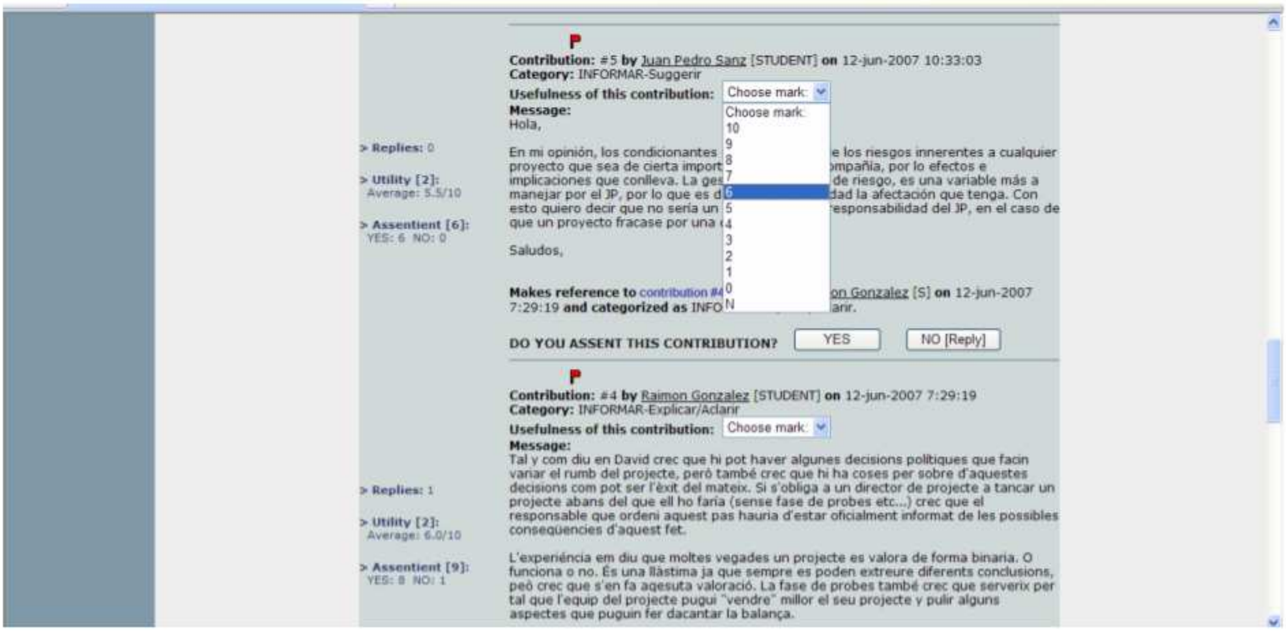

Figure 5.7: An example of a contribution to be assented and evaluated by a participant.

assessment (see Figure 5.7) and becomes very smoothly even in the case of large-size groups. This is achieved by first selecting the suitable mark of each contribution in a discussion thread and then by performing a single submission to assess all the contributions in the thread. The tutor assessment, along with the rest of indicators presented, are automatically and constantly processed by the system. For monitoring purposes, the system proposes an updated final mark of the progress of each student based on all the indicators presented (see Figure 5.8). These indicators are to be adjusted with appropriate weights by the tutor so as to reinforce certain aspects of the discussion and collaboration process according to the specific pedagogical objectives of the learning task.

Finally, for the sake of a rapid prompt of the awareness and feedback information to students and tutors, this research line has successfully managed to embed this information into the group activity in an efficient manner, even in real time. Indeed, the experience at the UOC has shown the need to monitor and evaluate real, long-term, complex, collaborative problem-solving situations. As a result, there is a strong need for powerful tools that record the large volume of interaction data generated during the group activity and can then be used to perform an efficient interaction analysis and knowledge extraction. Given the real needs of any online collaborative learning situation, in order to provide different types of awareness and feedback, it is required to capture all and each type of possible data that could result to a huge amount of information that is generated and gathered in data log files. Moreover, the need to make the analyzed information available in real time entails coming across with processing requirements beyond those of a single computer. To this end, several studies (for a detailed description of these studies, see Caballé et al., 2008a) have been conducted to show how a parallel approach can increase the efficiency of processing a large amount of interaction data and achieve an effective embedding of the appropriate knowledge extracted into collaborative learning practices.

Group formation
Planning
Discussion room
Repository
Voting system
Profile
Preferences
User directory
Change course
Start

You got NEWS:

FOLDER #1 - Proves (Santi Caballé [T]) 22/74 23/43

FOLDER #2 - Pact-Aprovació i definició (Pere Marine [T]) 3/354 18/340

[PERE] [TUTOR] [MGPI] [AULA 1]

FOLDER: #4 - 2on debat - Execució i tancament

Description: 2on debat - Execució i tancament de projectes
Created by: Pere Marine [TUTOR] on 01-jun-2007 14:25:21
FOLDER DATA:
N. contributions: 199 Quality of this folder: C+ Usefulness of this folder: 6.5/10 [615 Vots]

STUDENT STATISTICS

STUDENT POSITION		ACTIVITY				PASSIVITY		IMPACT		EFFECTIVITY	ASSESSMENT		
Pos.	Student	Total contributions	Proactivity	Reactivity	Support	Pending to read	Pending to evaluate	Particip. impact	Replies received	Assientment rebut	Peer assessment	Tutor assessment	FINAL MARK
[1]	Jaume Casadesús	9/199 4.5%	4.0/9 44%	5.0/9 55%	0.0/9 0%	90/190	190/190	3.5	5/9	28/28 100%	6.6/10 (28)	B	6.91
[2]	David Fernández	4/199 2.0%	3.0/4 75%	1.0/4 25%	0.0/4 0%	116/195	195/195	21.0	3/4	28/31 90%	6.7/10 (17)	B	6.56
[3]	David Recasens	4/199 2.0%	3.0/4 75%	1.0/4 25%	0.0/4 0%	182/195	195/195	24.5	0/4	33/33 100%	7.3/10 (11)	B	6.5
[4]	Sergio Acebes	18/199 9.0%	3.0/18 16%	15.0/18 83%	0.0/18 0%	9/181	10/181	-7.0	11/18	81/88 92%	6.6/10 (37)	B	6.44
[5]	Joan Ramon López	8/199 4.0%	1.0/8 12%	7.0/8 87%	0.0/8 0%	117/191	130/191	-3.5	4/8	27/29 93%	6.8/10 (24)	B	6.4
[6]	César Marin	8/199 4.0%	5.0/8 62%	3.0/8 37%	0.0/8 0%	146/191	191/191	13.5	2/8	41/41 100%	6.3/10 (26)	B	6.34
[7]	Vicent Moncho	6/199 3.0%	0.0/6 0%	6.0/6 100%	0.0/6 0%	33/193	162/193	-3.0	8/6	47/52 90%	5.8/10 (22)	B	6.28
[8]	Francisco José Garcia	9/199 4.5%	2.0/9 22%	7.0/9 77%	0.0/9 0%	124/190	190/190	14.0	10/9	67/75 89%	6.9/10 (36)	B	6.27
[9]	Luis José Salamó	12/199 6.0%	2.0/12 16%	10.0/12 83%	0.0/12 0%	74/187	69/187	8.0	7/12	17/17 100%	6.4/10 (28)	C+	6.26
[10]	Lidia Castillejo	3/199 1.5%	2.0/3 66%	1.0/3 33%	0.0/3 0%	193/196	196/196	-0.5	0/3	3/3 100%	6.5/10 (8)	B	6.22
[11]	Noemi Vázquez	5/199 2.5%	4.0/5 80%	1.0/5 20%	0.0/5 0%	168/194	194/194	2.0	3/5	13/13 100%	5.7/10 (15)	B	6.19
[12]	Lluis Martinez	20/199 10.1%	1.0/20 5%	19.0/20 95%	0.0/20 0%	23/179	179/179	-9.0	7/20	15/15 100%	6.3/10 (61)	C+	6.18
[13]	Iván Ruiz	6/199 3.0%	5.0/6 83%	1.0/6 16%	0.0/6 0%	191/193	193/193	18.5	0/6	19/19 100%	6.7/10 (15)	C+	6.13
[14]	Jordi Xargay	7/199 3.5%	1.0/7 14%	6.0/7 85%	0.0/7 0%	111/192	182/192	-3.0	3/7	58/61 95%	6.1/10 (29)	B	6.1
[15]	Aleix Cerecedo	7/199 3.5%	3.0/7 42%	4.0/7 57%	0.0/7 0%	143/192	192/192	13.5	3/7	57/62 91%	6.6/10 (27)	C+	6.03
[16]	Adoración Garcia	5/199 2.5%	1.0/5 20%	4.0/5 80%	0.0/5 0%	143/194	156/194	-2.0	6/5	14/16 87%	6.6/10 (20)	B	6.03
[17]	Joan Pons	7/199 3.5%	0.0/7 0%	7.0/7 100%	0.0/7 0%	95/192	102/192	-3.5	6/7	27/31 87%	6.6/10 (21)	C+	5.98
[18]	José Javier Doval	1/199 0.5%	1.0/1 100%	0.0/1 0%	0.0/1 0%	194/198	197/198	11.0	2/1	12/14 85%	5.8/10 (6)	B	5.98
[19]	Marc Berbel	4/199 2.0%	3.0/4 75%	1.0/4 25%	0.0/4 0%	190/195	195/195	10.0	3/4	11/12 91%	6.6/10 (16)	C+	5.95
[20]	Antonio Ponce	11/199 5.5%	3.0/11 27%	8.0/11 72%	0.0/11 0%	1/188	160/188	-4.0	10/11	56/67 83%	6.5/10 (27)	C+	5.95
[21]	Alexandre De Alcantara	6/199 3.0%	4.0/6 66%	2.0/6 33%	0.0/6 0%	179/193	193/193	9.0	4/6	21/23 91%	6.1/10 (24)	C+	5.82

Figure 5.8: Monitoring information provided to the tutor. Last column shows a numeric mark on the scale 0-10 for each student automatically generated and updated by the system. This final mark is based on all the indicators presented, which are adjusted with different weights. For the purpose of this specific discussion, the weights were set as follows: activity: 10%; passivity: 10%; impact: 20%; effectiveness: 10%; assessment: 50%.

5.2.3 Implementation issues

The DF system is hosted[3] by a server at the UOC. This prototype is currently working as a typical client-server Web-based application and evolving rapidly to be completed.

Taking advantage of the flexibility of the service-oriented approach forming its internals, different languages were used for the development of the DF's client and the server sides. Thus, on the one hand, PHP[4] resulted in a very suitable programming language to implement the web pages forming the user interface on the client side. On the other hand, the generic web-services supporting the business and data layers on the server side were implemented in Java as a powerful and experienced language featuring very well as to robustness, portability, ease of use and extensibility, which created an ideal context for the implementation on the server side. In order to interoperate between PHP and Java, the PEAR[5] framework was used.

5.2.4 Results and discussion on the effects in the learning experience

In order to evaluate the prototype of the DF and analyze its effects in the discussion process, 80 graduated students enrolled in the course Methodology and Management of Computer Science Projects during the last term were involved

[3]The DF's web site is found at: http://clpl.uoc.edu/df/ (Web site as of April 2008)
[4]PHP programming language is found at: http://www.php.net (Web site as of April 2008)
[5]The PHP Extension and Application Repository (PEAR) is found at http://pear.php.net (Web page as of April 2008)

Table 5.2: Main statistics extracted from the first class assignment using both discussion tools.

Statistics	**Standard tool**	**DF**
Number of students	40	40
Number of threads	48	44
Total of posts	95	351
Mean number (posts/thread)	M=1.9 SD=2.4	M=7.9 SD=5.0
Mean number (posts/student)	M=2.3 SD=1.9	M=8.7 SD=8.1
Mean number (words/contribution)	M=352 SD=139	M=286 SD=85
Tutor assessment (average, out of 10)	7.2	7.6

in this experience (Caballé et al., 2008b; Caballé et al., 2008h). Students were equally distributed into two classrooms and participated in the experience at the same time. Students from one classroom were required to use the standard asynchronous threaded discussion forum offered by the virtual campus of the UOC while the other group of students used the new DF outside the virtual campus to support the same discussions with the same rules during the same time (i.e., five weeks in all).

The whole experience consisted in two discussion assignments separated in time with very different goals and procedures so as to validate the flexibility of the approach. The first assignment in both groups lasted two weeks and consisted of discussing the same issue: *project management requirements vs. product requirements*. In this assignment, each student was required to start a discussion thread with posting a contribution on the issue in hand, which resulted in as many threads as students. At the end of the discussion, each student was asked to close his/her thread with an improved contribution on the issue according to what s/he had learnt in the discussion. During the discussion, any student could contribute in both the own and any other discussion thread as many times as needed, as well as start extra threads to provide new argumentations or approaches with regards to the issue addressed. The aim was to evaluate the effect of the discussion process in the acquisition of knowledge of each student by comparing the quality of each thread's first and last contribution posted by the same student.

A statistical analysis of the results in the first discussion comparing both the standard and the DF tools is shown in Table 5.2. Despite the standard tool generated more threads, most of them were actually empty (i.e. just 8 threads were contributed with more than 1 post vs. 42 threads in the DF). Moreover, the SD statistic for the posts/thread mean appears to be high in the DF, which proves the heterogeneity of the discussion involving threads of very different length. Note the very high SD statistic in the posts/students mean due to a single outlier, without which SD is 6.3. Finally, quality statistics are shown in terms of the number of words per contribution and the tutor assessment on the content. The higher number of number of words in the standards tool is due to the lack of discussion as most of threads were just started with a long opening contribution as a problem statement. On the other hand, the DF generated actual discussion and as a result the contributions became highly structured and specific. The tutor assessment row refers to content quality of all the contributions on average.

The qualitative evaluation of this first discussion was addressed by both examining those discussion threads that contained enough discussion (i.e. more than 7 posts) and checking whether the student who was in charge of each thread

Table 5.3: Main statistics of the second class assignment using both discussion tools.

Statistics	**Standard tool**	**DF**
Number of students	40	40
Number of threads	43	21
Total of posts	71	199
Mean number (posts/thread)	M=1.6 SD=0.4	M=9.4 SD=3.2
Mean number (posts/student)	M=1.7 SD=1.1	M=4.9 SD=4.1
Mean number (words/contribution)	M=421 SD=127	M=310 SD=96
Tutor assessment (average, out of 10)	8.1	7.5

had posted both a start and close contribution on the same issue. The results on the DF showed that, in 28 threads fulfilling these requirements, 32% of students had improved their qualitative mark through the discussion in their threads, 68% kept the same mark, and no mark had dropped. On the other hand, no results were extracted from the discussion using the standard tool as it was poorly contributed; just 8 threads showed some discussion but only 4 had more than 7 posts.

The second assignment in both groups was held at the end of the same academic term, one month after finishing the previous one and lasted for three weeks. It consisted of discussing the stage of closing a software project. The procedure was the following: students were free to open zero, one or several discussion threads where they proposed specific objectives, activities, and processes needed to appropriately close a software project. Hence, in this discussion there was no requirement to open a discussion thread and all students could participate in the discussion threads at convenience. At the end of the discussion, those students who had opened a discussion thread were asked to close it by sending a contribution that summarized and concluded the main points arisen in the thread.

The statistical analysis of the results extracted from the second discussion comparing both the standard and the DF tools are shown in Table 5.14. In comparison to the previous assignment, there is a decrease in the number of contributions in both groups, which may be explained by two reasons: even though the number of potential participants was the same as the previous discussion, 40% in each group had already made the decision to give up the course before this second discussion started and as a result most of them did not pay attention nor contribute to the discussion. Moreover, the participation was not a requirement in this assignment and consequently some students chose not to participate.

The qualitative evaluation of the second discussion was addressed in a similar way as the previous one. Despite the standard tool rated high in the number of threads, just one of them provided real discussion (i.e., more than 7 posts). Hence, it could be stated that no discussion was achieved using the standard tool. On the other hand, the DF performed much better providing actual discussion in 16 out of 21 threads achieved.

The mean number of words per contribution in the standard tool also rated higher than the DF in this second experience. This confirms the effects of the inherent structure and richness provided to the discussion process by the DF whereas the standard tool promotes large monolithic one-sided points of view. Finally, the standard tool achieved a higher average mark on the qualitative content of the contributions. It could be argued that most of participants of the standard tool were good students, whose first and only contribution to a thread was fine. However, the lack of discussion missed many important aspects, such as reactive participation behavior and peer involvement skills, which are fundamental to achieve a successful discussion process. All these aspects are to be combined with the evaluation of the qualitative content to form the final assessment of the collaborative learning activity.

Table 5.13 shows the results of a structured and qualitative report conducted at the end of the first discussion addressed to the DF' users who were also asked to compare it to the standard well-known tool they had already used in previous courses at the UOC. Participants in general showed an optimistic attitude toward this experience, which sheds more light on the positive effects in the learning experience. However, participants also reported technical problems while participating, such as the lack of fault-tolerance, performance and concurrency. This undoubtedly undermined the overall benefits achieved. In order to alleviate this problem, the second discussion was supported by an innovative distributed approach so as to meet certain non-functional requirements that were not considered in the first experience, such as fault-tolerance. The results of the second discussion from this standpoint are reported later on in this chapter in a related section.

Finally, in order to evaluate the reliability of the automatic assessment approach in both assignments, the tutor supervising the discussions supported by the DF was required to both (1) submit a precise assessment on content quality of every contribution posted, which was presented to students as feedback information (see Figure 5.3 and section 2.2 in Chapter 2 for further information on tutor assessment) and (2) evaluate students' performance manually by the tutor by filling out a spreadsheet that helped score each student's participation according to both the content quality of each of his/her contribution and the purpose and context where the contribution took place (e.g., whether it was a new argumentation or a reply, brought interesting opportunities for further discussion, it was just a greeting-type post, etc.). This second evaluation task could be complemented with extra information on individual behavior in the discussion added by the tutor according to his knowledge and experience in this type of class assignment.

The ultimate aim of this double evaluation process was to compare the manual evaluation performed by the tutor to the semi-automatic assessment process provided by the system. To this end, each evaluation process resulted in proposing both a final mark for each student and a position list where all students were ranked according to his/her final mark (see first and last columns in the feedback information depicted in Figure 5.8. In the semi-automatic evaluation, on the one hand, the system addressed four indicators, namely, activity, passivity, impact and effectiveness, becoming 50% of automatic evaluation. The rest of the evaluation came from the quality indicator only, which was addressed by the tutor who was in charge of assessing the contributions' content quality (40%), and the peers who assessed the usefulness of others' contributions on average (see also Figure 8 for further information). Please note that these percentages may vary according to the type of the discussion and they can be adjusted by the tutor. On the other hand, the manual evaluation process was carried out entirely by the tutor and followed the same assessment procedure as that performed while using the standard discussion tool of the UOC.

The results of the automatic assessment were very promising since the tutor in charge of the DF agreed with the final marks proposed by the system in more than 75% of cases. 31 out of 40 students in the DF's rank matched the same position as in the rank appeared in the tutor's spreadsheet. In addition, the tutor reported the promising benefits from the

Table 5.4: Excerpt of a questionnaire's results on both the DF and the standard tool to support the discussion process.

Selected questions	**Average of structured responses (0 - 5)**	**Excerpt of students' comments**
Assess the DF as a collaborative tool	3	”Apart from some technical problems, the DF fulfilled my expectations”
Evaluate how the DF fostered your active participation	4	”The statistical data and quality assessment displayed influenced my participation”
Did the DF help you acquire knowledge on the debate's issue?	4	”The standard tool is a chaos for large debates (...) DF's discussion rooms eased me the debate a lot”
Compare the DF to the campus' standard discussion tool	4	”The DF should be used to support debates in other courses”

DF in the monitoring process on the discussion since this new tool alleviates tutors and moderators from the tedious work of tracking and evaluating the discussion's dynamics and outcomes manually. On the other hand, a clear inconsistency was identified since all final marks proposed by the system scored 1.1 points on average lower that those proposed by the tutor, thus showing the need to weight the indicators in the DF more objectively. In overall, these results are not conclusive but they encourage us to undertake more experimentation and especially validation processes on the automatic assessment approach.

To sum up, the overall experience is very promising for enhancing knowledge management that contributes to the improvement of the discussion process in virtual collaborative learning environments. Even though the results of this experience are not conclusive due to its exploratory nature, from the analysis of the results it has been proved to promise significant benefits for students in the context of project-based learning, and in education in general.

On the other hand, the centralized approach in terms of a single physical node supporting the whole discussion process brought considerable repercussions on the learning experience, such as lack of performance, scalability, fault-tolerance, and interoperability (see section 4.2 in Chapter 4 and Caballé et al., 2008e for more information). To this end, later on in this chapter, it is reported the experience of providing distributed infrastructure to the DF prototype in order to meet the mentioned non-functional requirements that may influence the learning process a great deal (Caballé et al., 2007c).

5.3 CoLPE: support for communities of learning practice

This section reports on a experience of using an innovative technology-enhanced learning tool to support a real community of formal learning practice (Caballé and Feldman, 2008f). First, the underlying groupware platform, called CoPE, is introduced that provides the essential functional support for democratic groupware. Then, the main guidelines for the requirements and design of this application are described. As part of the design, specific action types are proposed that promote meaningful contributions to be used to analyze learners' interactions in terms of performance and the particular skills exhibited during interaction. The aim is to extract relevant knowledge in order to provide learners and tutors with efficient awareness, feedback, and monitoring as regards learners' performance and collaboration. Finally, this tool is employed in a real on-line learning environment to support a collaborative activity based on an asynchronous discussion.

The experience and the evaluation results of using this application are reported, showing promising opportunities to support the formal and also informal discussion processes occurring in current communities of learning practice.

5.3.1 Motivation

Over the last several years, collaborative e-Learning needs have been evolving with more and more demanding pedagogical and technological requirements (Pankatrius et al., 2003). Modern pedagogical approaches targeting formal education include advanced learning techniques based on some form of collaborative consensus-building mechanism, such as learning by discussion and problem-based learning (Dillenbourg, 1999). To this end, a great deal of software packages in the form of Learning Management Systems (LMS) has recently appeared in the marketplace to support those communities of learning practice formed during the formal learning process, which typically involve all students in a classroom. These tools enable the management of educational content and also integrate tools that support many of the groupware needs, such as e-mail, discussion forums, chat, virtual classrooms, and so on (Caballé et al., 2007b).

On the other hand, informal collaborative learning typically involves a small number of students who meet each other informally after classes in small study groups to carry out specific learning activities assigned during the formal learning process. These groups of people also form communities of learning practice where an important part of both individual and group learning process takes place and whose members are often separated geographically and have the need to meet asynchronously.

In all cases, collectives of students who are separated by geography and/or time form communities of learning practice where an important part of both individual and group learning process takes place asynchronously. However, the lack of suitable and available groupware applications makes it difficult for these groups of learners to collaborate and achieve their specific learning goals. In addition, current collaborative learning applications and sophisticated learning management systems do not conveniently address the support to learning groups who are chiefly formed by non-technical people and who lack of the necessary resources to acquire such systems, especially in informal learning.

In particular, in on-line collaborative learning environments, the discussion process forms an important social task where participants can think about the activity being performed, collaborate with each other through the exchange of ideas arising, propose new resolution mechanisms, and justify and refine their own contributions and thus acquire new knowledge (Salomon, 1993). The lack of technological support for democratic decision-making mechanisms is however a main handicap to both achieve a consensus in a discussion process by means of voting and substitute the central authority of knowledge in small study groups. Furthermore, current collaborative applications provide poor support for the representation and analysis of group activity interaction as an essential feature to sustain a collaborative learning discussion, in terms of coaching, monitoring, and evaluation (Zumbach et al., 2003; Daradoumis et al., 2006). A large amount of information is generated from the actions performed by the participants during the discussion process, which includes complex issues of the collaborative work and learning process (e.g., group well-being as well as self, peer and group activity evaluation). This information is then used in extracting and providing effective knowledge on interaction behavior to adequately regulate the learning process as well as to enhance learning group participation by means of providing appropriate awareness and feedback.

In this section, these entire approaches are taken one step further by introducing a new collaborative learning tool called Communities of Learning Practice Environment (CoLPE), as a result of a research effort of both the Open University of Catalonia and the International Computer Science Institute[6]. CoLPE was developed to support and enhance the discussion process encountered in many on-line courses and also in those informal study groups in the form of online discussions. This system implements many of the approaches described so far and the first results drawn from real

[6]The International Computer Science Institute (ICSI) is a leading research center in computer science located in Berkeley, CA, USA. ICSI can be reached at: http://www.icsi.berkeley.edu (Web page as of April 2008)

collaborative learning show very promising benefits for students in a real context of learning and in education in general.

To this end, first, it is presented an existing groupware system called CoPE developed by the same research group that provides informal support to collaborative work. Then, the main requirements that guided the development of CoLPE are described by means of extending CoPE to the learning domain and incorporating essential functionalities regarding the management of information and knowledge about group activity. The experience and the evaluation results of using this application in a real learning context are reported later on in sub section of results.

5.3.2 CoPE: Democratic support for collaborative work

CoPE[7] (Feldman et al., 2006; Thaw et al., 2008) is a web-based collaborative system aiming at providing formal and informal cooperative work over the Internet to non-technical people or those who lack of the necessary resources to acquire such systems. As such, CoPE provides most of the functionality expected from an asynchronous Computer-Supported Collaborative Work (CSCW) (Bentley et al., 1997) application, such as information management and communication facilities.

CoPE is designed to enable a specific type of collaboration; a subset of CSCW that has not been adequately addressed so far. Specifically, this involves sets of individuals who share a need or desire to engage in collaborative production. The object of this production is something that can be codified in documents. CoPE is targeted to individuals who do not already have a formal workflow for this collaboration or who are seeking to improve upon inefficient workflows. CoPE also envisions enabling collaboration among individuals who are part of organizations with formal collaboration mechanisms, but whose mechanisms are limited to intra-organization collaboration. Finally, CoPE is designed to enable collaboration, not management, and thus envisions ”democratic” collaboration.

There are many examples of sets of individuals around the world who have a need or desire to collaborate but lack the resources, knowledge, or institutions to do so. Consider, for example, public school teachers, social workers, and community action groups (where the group and its peer groups are the ”individual”). Often these individuals are separated by geography and/or time and can be too distant from one another to organize face-to-face meetings. They also could be unable to meet due to scheduling constraints or differing work hours. Such individuals may already be part of existing organizations but the ”peers” with whom they wish to collaborate are in different organizations. CoPE is especially targeted to these individuals and organizations who lack substantial technical expertise or the resources to acquire such expertise.

CoPE was developed for the needs of a certain type of user forming the CoPE User Community. The system interface design makes assumptions based on the characteristics of such users. We call this type of user the ”General User”. The following assumptions motivate this definition:

- users do not have specialized (information/computing) technical skills,
- users possess a basic skill set for computer and Internet usage,

[7] The CoPE’s web site is found at: http://cope.icsi.berkeley.edu (Web site as of April 2008)

- users posses the ability to learn a new (information/computing) skill set of this same basic technical level,
- users are willing to learn a new (information/computing) skill set of this same basic technical level,
- users do not already share a sophisticated and/or long-used method for electronic collaboration.

There are several features and mechanisms of the implemented CoPE system that support collaborative work and in particular group discussions:

- hierarchical threaded discussion of documents to serve as a core for group consideration of material of any kind, which can include arbitrary additional material where a coordinator typically posts a document for discussion and also intervenes in the ongoing dialog when appropriate.
- support for the production of joint projects by subgroups of participants by easily setting up subgroups so that the work of each group is kept private from the others, but is visible to the coordinator.
- allow the coordinator of a CoPE site to customize much of the form and content of the material without programming and a range of choices on discussion and voting methods are provided enabling coordinators without IT expertise to customize their discussion environments.

CoPE is built by modifying and taking advantage of Plone/Zope's (Aspeli, 2007) powerful content management capabilities, such as information management, document workflow, and so on. CoPE modifies Plone appropriately to achieve the desired functionality.

5.3.3 CoLPE Development

The extension of CoPE to learning is called Communities of Learning Practice Environment (CoLPE) (Caballé and Feldman, 2008c; Caballé et al., 2008f) which relies heavily on CoPE, and in turn on Plone, for most of the mentioned functionality that combines CSCW and collaborative learning paradigms. In addition, specific behavior has to be aggregated to facilitate both the construction of knowledge among learners and the development of cognitive-acquisition skills, such as problem-solving abilities as well as the provision of an adequate multi-support framework so that tutors and peers can provide a suitable scaffolding when needed; these are key aspects that distinguish CSCL from CSCW. CoLPE pursue theses objectives by means of seeing discussion as a medium through which the building and distribution of skills and knowledge is effected.

This subsection presents the collaborative learning requirements that motivated the CoLPE development and the main guidelines that guided its design. The ultimate aim is to provide full support to both formal and informal learning groups by means of the collaborative discussion process.

5.3.3.1 General requirements and pedagogical background

CoLPE's requirements[8] include support for the essential types of generic contributions found in any discussion process, namely specification, elaboration and consensus (Salomón, 1993). Specification occurs during the initial stage of the process carried out by the tutor or group coordinator who contributes by defining the group activity and its objectives (i.e. statement of the problem) and the way to structure the group activity in sub-activities. Elaboration refers to the contributions of participants (mostly students) in which a proposal, idea or plan to reach a solution is presented. The other participants can elaborate on this proposal through different types of participation such as questions, comments, explanations and agree/disagree statements. Finally, when a correct proposal of solution is achieved, the consensus mechanisms take part in its approval (this includes different consensus models such as voting); when a solution is accepted the discussion terminates.

A fundamental requirement to sustain collaborative learning applications is the representation and analysis of group activity interaction to facilitate coaching and evaluation (Dillenbourg, 1999) as well as awareness and feedback about what is happening during the collaboration. To this end, in extending CoPE to e-Learning a primary requirement is management and provision of information and knowledge about group activity. The ultimate goal is to enhance and improve group activity by constantly keeping users aware of what is going on in the system (e.g. others' contributions, new documents created, etc.), In addition, monitoring participants' performance allows tutors to identify problems that participants may encounter during the assignments. These findings can then be used to provide both real-time and asynchronous support to students (i.e., help students who are not able to accomplish the tasks on their own).

Finally, in a discussion process, participants perform a role according to their profile (e.g., coordinator, member, guest, etc.), have personal preferences (e.g., language) and set up environment features (e.g., sound or visual effects, text or voice warnings, etc.) according to their personal characteristics. Participant needs are not static and they evolve as the discussion moves forward.

5.3.3.2 The design of the application

CoLPE design[9] aims at providing specific support to the essential types of generic contributions in a discussion process identified in the requirements, namely specification, elaboration and consensus. In CoLPE, these different types of generic contributions are managed by the three essential aspects existing in any collaborative learning application (i.e., coordination, collaboration and communication) (Ochoa et al., 2002; Caballé et al., 2004). Coordination involves the organization of groups to accomplish important objectives to perform a discussion, such as workspace organization, group structure and planning. Collaboration lets group members share any kind of resources while communication represents the basis of the whole discussion process since it enables coordination and collaboration to be achieved by providing them with low-level communication support. Based on these three areas of cooperation, the main guidelines in designing the generic

[8] The complete requirements of CoLPE and their analysis are found at: http://clpl.uoc.edu/docs/CoLPESpecification.pdf (Web page as of April 2008)

[9] The complete design of CoLPE is found at: http://clpl.uoc.edu/docs/CoLPEDesign.pdf (Web page as of April 2008)

types of contribution is as follows:

- The specification phase is mainly based on coordination which involves the organization of groups such as workspace organization and group structure and planning, so as to accomplish group objectives.

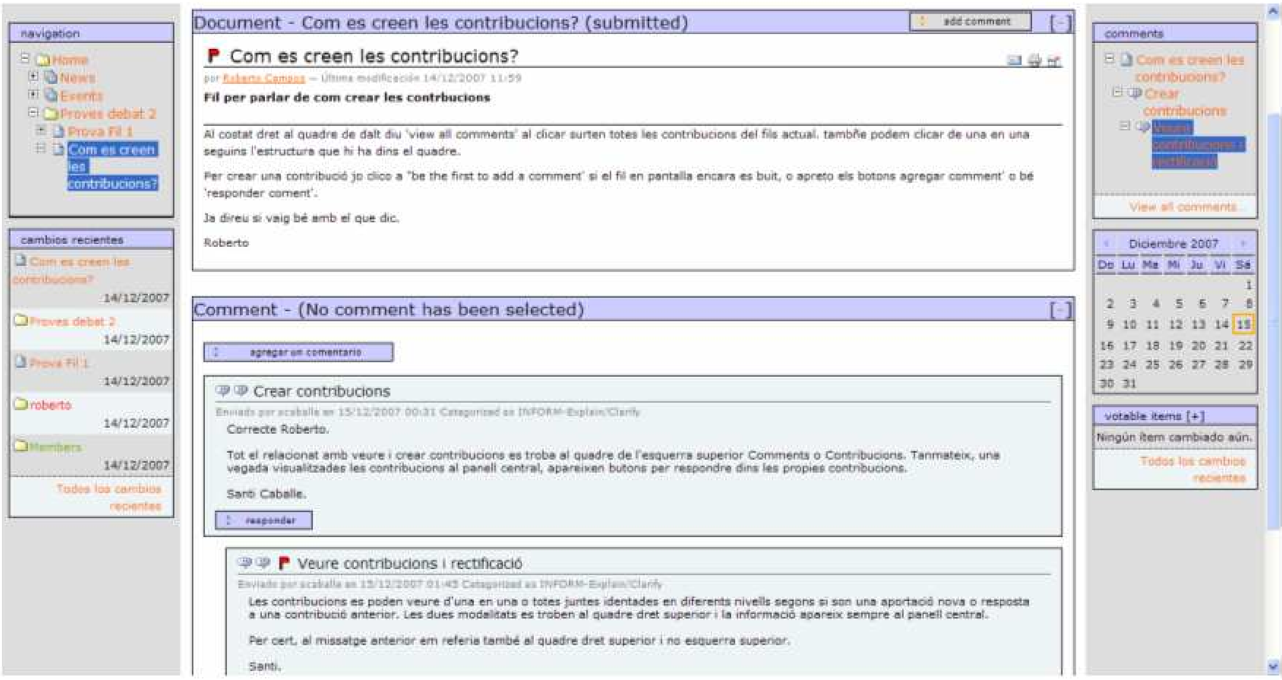

Figure 5.9: Discussion thread formed by the head of the thread and the follow-up comments. Red flags provide feedback at thread level informing where the news is.

- Elaboration is the main phase in the discussion, which relies on both collaboration and communication allowing students to share any kind of resources (e.g., participation spaces, documents, etc.) as well as exchange ideas by posting messages to a discussion space. To this end, this phase is mainly structured in CoLPE by means of folders, which hold the discussion threads and other subfolders forming the whole discussion as a learning assignment or activity. A discussion thread in turn holds a document or text page, which will head the rest of the comments of the same thread. A subfolder may contain others in order to organize the workspace more effectively or for the purpose of storing additional resources (see Figure 5.9).

  During the elaboration phase, a key issue in CoLPE is that before a participant sends a new contribution to a discussion thread, this contribution is categorized using a predefined list of labels or categories, such as request for information, opinion, clarification, elaboration, etc.; inform in terms of extension, suggestion, explanation, justification, illustration, etc.; problem, which may be found as statement, solution, etc; greetings, motivation, among others (see Figure 5.10 and also Table 5.1) for a complete list of labels). The purpose of these categories is to classify the intention of the contribution. Not all categories are always made available since depending on where the discussion is found just a subset of them are made available. These categories represent the information source to eventually present complex feedback to users in terms of participation impact and user profile (see further in this

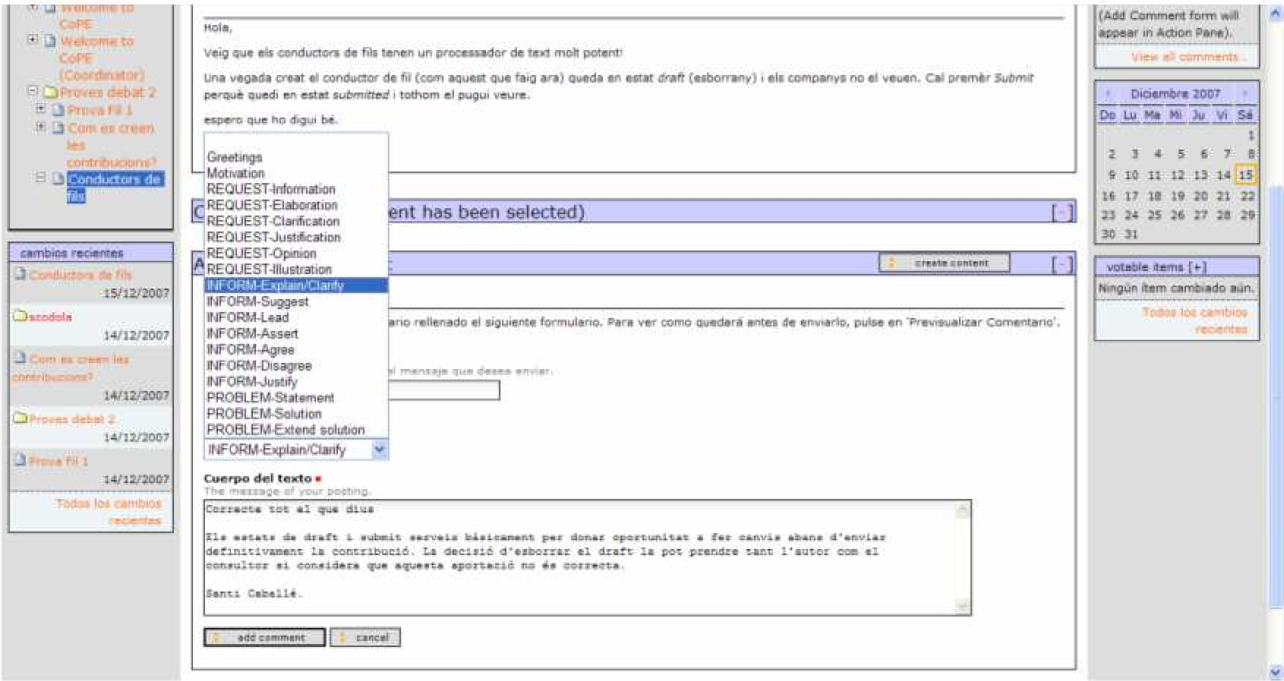

Figure 5.10: A list of tags to qualify a contribution.

section for details).

Depending on the pedagogical model and objectives pursuit in the discussion, a discussion thread may start by submitting a proposal, a solution of a problem, etc., which is to be later on discussed by the participants by means of sending contributions to the thread. Eventually, based on the cognition level achieved during the discussion, participants may vote on the initial proposal submitted so as to approve/disapprove it. On approval the proposal may be archived while on disapproval it may be also revised and resubmitted to be discussed again. Therefore, a discussion thread follows a workflow with several states, from draft to approval or rejection (see Feldman et al. (2006) for more information). The functionality is available to the tutors who can manually change the state of the thread.

- The consensus phase in the discussion process is also based on collaboration by which a voting system is shared by the group members to choose the best proposal arisen during the discussion. To this end, several voting modes are available in CoLPE to meet different consensus needs (see Figure 5.11.

In order to equip CoLPE with appropriate knowledge management of the users' interaction data analysis, the generic, reusable service-oriented, component-based capabilities of the CLPL platform was used (CLPL) (Caballé et al., 2007e).

This platform was specially used to give support for the interaction data analysis process and the presentation of the knowledge extracted (see section 2.3 in Chapter 2 for an overview of this process). I particular, the CLPL's capability of managing the system's log files and the statistical analysis event information as well as the management and maintenance of the knowledge extracted by that analysis. In addition, the design took advantage of the presentation of the knowledge

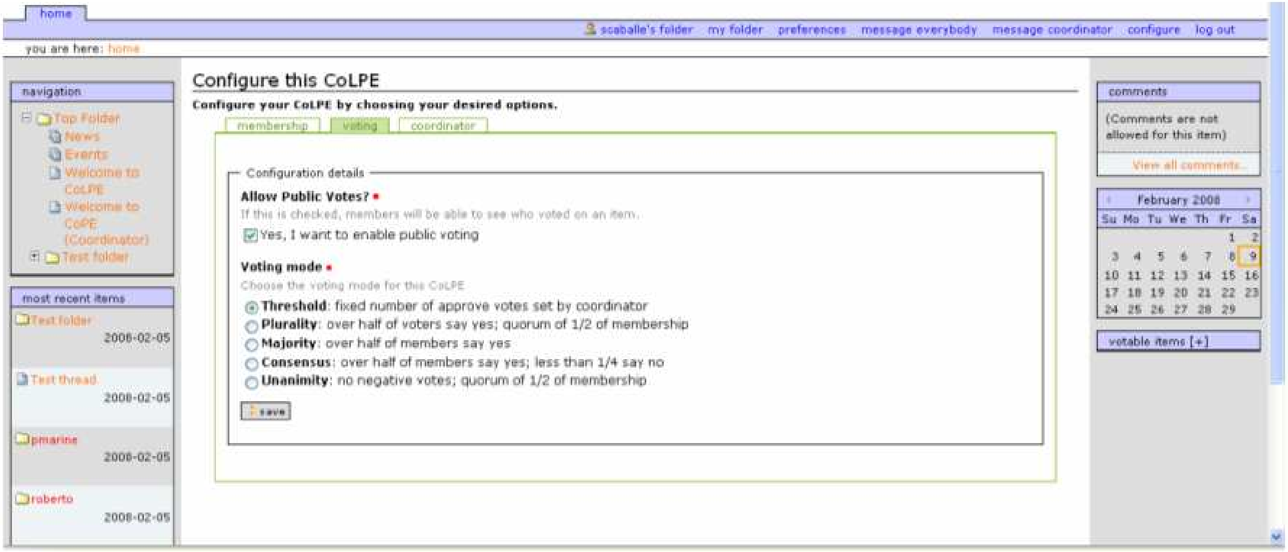

Figure 5.11: Several voting modes available.

generated to users in terms of immediate awareness (see Figure 5.8 and constant feedback (see Figure 5.12) of what is going on in the system.

Finally, personal features of the discussion group participants (their role, collaboration preferences and so on) were taken into account and a user and group model were designed so as to allow participants to add new services as their needs evolve as the discussion moves forward. These entire user features were included in CoLPE by means of the the CSCL User Management component through the user profile management subsystem, providing solid support for building and maintaining the user and group model.

Therefore, CoLPE supports a complete discussion process through the realization of three generic contribution types and an open user and group model. Furthermore, this application constitutes a valuable resource to improve essential features of a discussion process such as awareness of participant contributions and enhance the abilities of users by increasing their knowledge of each other in terms of motivation, interaction behavior and so on.

5.3.4 Implementation technology

Presently, CoLPE prototype is hosted[10] by a server at the UOC and running a client-server application. Unlike the DF, which was built entirely from the CLPL, CoLPE prototype was originally an existing system to support collaborative work (see CoPE system above) and extended to the learning domain by means of the CLPL. This subsection sheds light on the internals of CoLPE from the technology standpoint.

Advanced frameworks in the form of web application servers have appeared in the marketplace for the easy Web development. One important reason that justifies their existence is to make the logical part of the applications independent from the presentation to users (Caballé et al., 2008g). Indeed, this is the most important decision while implementing a

[10]CoLPE's web site is found at: http://clpl.uoc.edu:8080/colpetest (Web site as of April 2008)

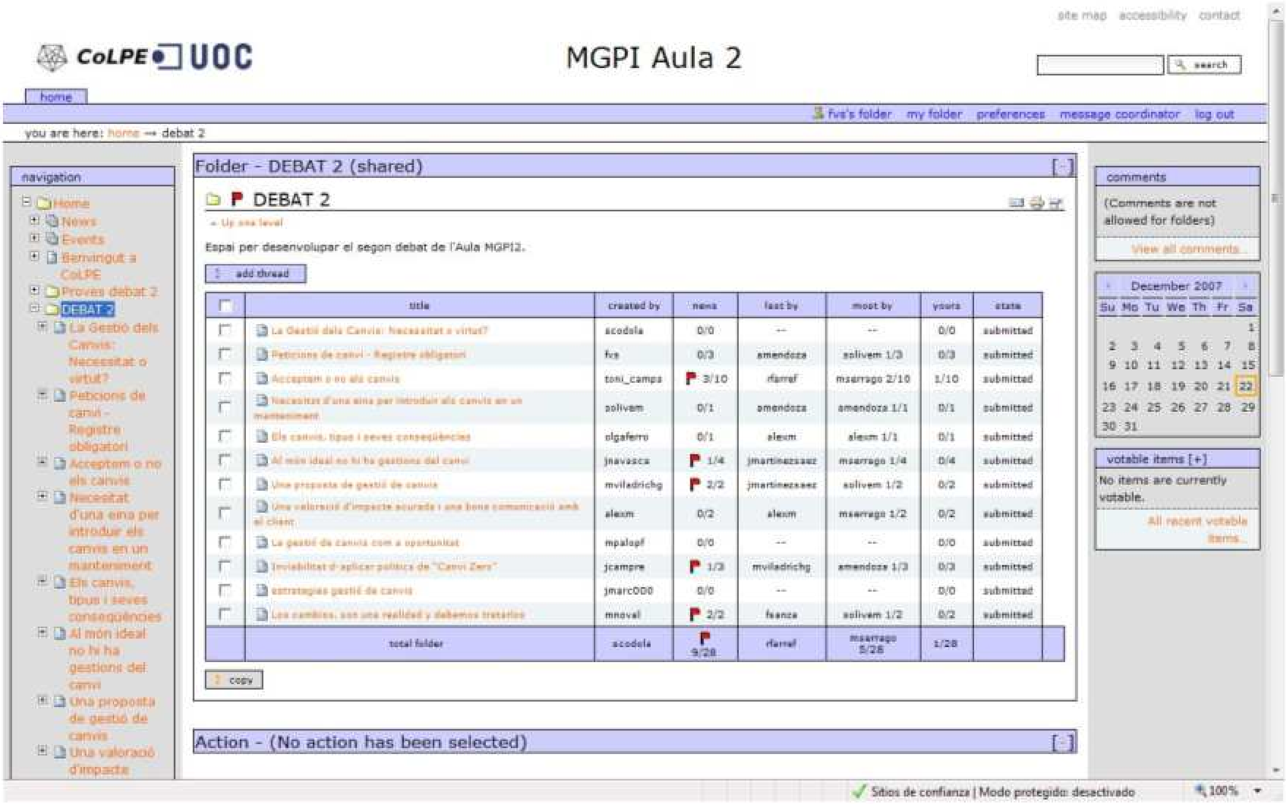

Figure 5.12: Partial feedback at folder level presented to all participants. The *most by* and *yours* indicators allow students to compare their own quantitative performance to the rest of the contributors.

software that was designed according to the MVC design pattern given that it makes it possible to distinguish the different roles involved in the implementation and thus the graphical designers can update the user interface without depending on the code programmers.

CoLPE is supported by a main representative of web application servers called *Zope*, which is open-source, object-oriented and python-based (Latteier et al., 2007). CoLPE takes advantage of Zope's object-oriented internal ad hoc database called Zope Database (ZODB). ZODB provides web programmers with advanced storage capabilities on Zope's objects without having to be worried about database issues, so the persistence management in Zope becomes completely transparent to them. Moreover, Zope comes with plenty of database adapters to connect Zope to any external database existing in the organization where Zope plans to be involved as well as the large Zope's API allows programmers to deal with all Zope's objects directly from scripts. Finally, a great deal of web programming and administration work can be easily done by using the powerful web-based Zope Management Interface (ZMI) which is immediately propagated to CoLPE.

Zope is built with Python[11]. There exist many projects aiming at providing support for developing web services with Python. Major effort is Zolera SOAP Infrastructure (ZSI)[12]. CoLPE takes great advantage of ZSI to extensively use the CLPL by calling its Web-services from the Zope's python-implemented internals.

[11]Python programming language is found at: http://www.python.org (Web page as of April 2008)
[12]ZSI is a SourceForge project found at: http://pywebsvcs.sourceforge.net/ (Web page as of April 2008)

The flexibility of the CLPL's Web-services was essential in interfacing Zope's functionalities into the CLPL environment and thus providing seamlessly computational learning support. This flexibility is two-fold. First, fully interoperability between Zope's Python and CLPL's Web-services coded in Java. Second, high level of abstraction of the CLPL's generic Web-services, which could be reused and appropriately filled in and coded with the specific functionality required by CoLPE. This independence between programming languages allowed for the use of Java to code the Web-services and take advantage of this powerful and experienced language for the implementation of the logics and information layers. .

On the other hand, a Web Content Management System (Web CMS) are software systems primarily used for Web content management among a potentially large number of contributors. The main purpose of a Web CMS is to make available a great deal of electronic documents and Web content over the Internet to a large amount of people who share and enrich such contents in a non-proprietary fashion. Some systems include interesting features such as the concept of workflow with the idea of moving web content or electronic documents so as to create and update and make decisions on a content collaboratively.

CoLPE relies heavily on one of the main exponent of Web-based open-source CMS called Plone (McKay, 2006), which is built on top of the Zope web application server. Plone provides advances CMS capabilities, such as information management, document workflow, groupware tools, and so on. The main advantage of Plone is to be extensible so as to fit specific needs in the CMS domain. For further information on Plone, see McKay (2006). CoLPE is built by modifying Plone and in turn Zope. Plone provides CoLPE with most of the content management technology needed, such as database storage, user authentication, and workflow, as well as presentation code, which is written in HTML and CSS sheet styles. CoLPE adds and modifies code in Plone to achieve desired functionality.

5.3.5 Evaluation and results

In order to evaluate CoLPE's prototype and analyze its effects in the learning experience, and in particular the discussion process, the real on-line learning context of the Open University of Catalonia was used. 43 graduate students enrolled in the course Methodology and Management of Computer Science Projects were involved in this experience.

5.3.5.1 Experiment procedure

The experience consisted of a discussion assignment, with the aim of discussing how a project manager can deal with the problem of changing the requirements of software projects which are already in advanced phases of their development because of demanding and urgent needs of the clients. The assignment title was: ”Change management: necessity or virtue?”.

The procedure was the following: students were free to open zero, one or several discussion threads (i.e., head of threads) where they proposed strategies, ideas, etc., to appropriately deal with the topic of the discussion. During the discussion, any student could contribute in both his own and any other discussion thread as many times as needed, as well as start extra threads to provide new arguments or approaches with regards to the issue addressed. The only requirement

Table 5.5: Basic statistics about participation.

Statistics	CoLPE
Number of students	43
Number of active students	26
Number of heads of threads	17
Number of comments in threads	93
Total of posts	110
Mean number (posts/thread)	M=6.4 SD=4.5
Mean number (posts/student)	M=4.2 SD=3.8

was to make at least one post to either a head of thread or a comment.

5.3.5.2 Results and analysis

The results of this experiment are provided by means of statistical analysis. A structured and qualitative report was also conducted at the end of the discussion addressed to all participants who were asked to both assess the prototype and compare it to the standard well-known discussion tool they had already used in previous courses at the UOC.

A statistical analysis of the results of the discussion is shown in Table 5.14. Note that the discussion took place at the end of the course and even though the number of potential participants was 43 (i.e., students enrolled in the course), roughly 40% of them had already made the decision to give up before the assignment started and as a result they did not pay attention nor contribute to the discussion. So, the number of active participants who participated in the discussion actively or passively was 26.

From the results of Table 5.14, the SD statistic for the posts/thread mean appears to be high, which shows the heterogeneity of the discussion involving threads of very different length and also that actual discussion was generated and as a result the contributions became highly structured and specific. In addition, the posts/student mean rates high (the requirement was 1 post per student) and shows a general interest in the discussion.

On the other hand, the SD statistics for posts/student is also high meaning that some students participated a lot (more than 10 posts) while a few tried to fulfill the assignment's requirement and provided single, monolithic point of view. It could be argued that at the end of the course students lack time, though more experimentation have to be undertaken to confirm these results.

Table 5.15 shows the most frequent categories used to tag the contributions. Although the choice of the category appears to be mostly correct, they could indeed be more precise. The permanent availability of all possible categories did not help participants to choose carefully. In future iterations, only those categories which are appropriate (i.e., make sense) at a certain point of the discussion will be shown, thus facilitating the choice a great deal.

Table 5.7 shows the results of a structured and qualitative report conducted at the end of the discussion addressed to the CoLPE users who were also asked to compare it to the standard well-known tool they had already used in previous courses at the UOC. Despite participants were excited with this experience, this report also shows the technical problems faced due to the server's poor performance where CoLPE was running, which was unable to conveniently handle both the

Table 5.6: Distribution of the tagged contributions.

Exchange actions	**Contribution categories**	**# Tagged contributions**
support	Greeting	3
	Motivation	0
request	REQUEST-Information	
	REQUEST-Elaboration	0
	REQUEST-Clarification	3
	REQUEST-Justification	0
	REQUEST-Opinion	20
	REQUEST-Illustration	0
inform	INFORM-Extend	17
	INFORM-Lead	0
	INFORM-Suggest	8
	INFORM-Elaboration	0
	INFORM-Explain/Clarification	17
	INFORM-Justify	1
	INFORM-State	0
	INFORM-Agree	21
	INFORM-Disagree	6
set-up-an-issue	PROBLEM-Statement	16
provide-solution	PROBLEM-Solution	1
consent-solution	PROBLEM-Extend solution	0

demanding hardware requirements of Zope server and the participants' concurrency (Caballé et al., 2008e). This problem is explored and handled in the next section.

In overall, this is a promising approach for enhancing communities of learning practice by means of an innovative tool that contributes to the improvement of the discussion process occurring in both formal and informal collaborative learning settings. Indeed, the analysis of the results promise significant benefits for students in the context of project-based learning, and in education in general.

On the other hand, more powerful hardware is needed so as to overcome the poor server's performance issue. To this end, next section reports the experience of adding distributed infrastructure to the CoLPE prototype in order to meet other important non-functional requirements that influence the learning process a great deal (see section 4.2 in Chapter 4 and Caballé et al., 2007g), such as scalability, fault-tolerance, and interoperability. For instance, the gain in fault-tolerance might help enhance the effectiveness of complex collaborative learning processes (e.g., by avoiding a central point of failure).

5.4 Adding distributed infrastructure to meet non-functional requirements

This section summarizes and reports from a different perspective on the same experiences of using the Discussion Forum (DF) and CoLPE as those previously described. Even though those learning experiences were successful from the pedagogical standpoint, many inconvenience arose in terms of technical issues that impacted negatively on the whole experience. Indeed, as discussed in Chapter 4 (see section 4.2), certain non-functional requirements are especially frustrating when they are not fulfilled appropriately during any collaborative activity. These requirements include fault-tolerance,

Table 5.7: Excerpt of the questionnaire filled out by the students.

Selected questions	**Average of structured responses (0 - 5)**	**Excerpt of students' comments**
Asses CoLPE as a collaborative tool	4	”Despite technical problems with the server I found CoLPE very useful due to the distribution of posts into threads and also be aware of where the news was”
Evaluate how CoLPE fostered your active participation	2	”I liked the categorization as it helped me understand others' contributions and reply being more confident on my contribution”
Did CoLPE help you acquire knowledge on the debate's issue?	2	””The notification of news was useful to be aware what was happening”
Compare CoLPE to the campus' standard discussion tool	3	”CoLPE is more suitable to support this type of discussion than the UOC's forum” ”Certain functions are missing in CoLPE: subscription to your thread, advanced search function, etc.”

Table 5.8: Main statistics extracted from the debate using both discussion tools.

Statistics	**Standard tool**	**DF**
Number of students	40	40
Number of threads	57	65
Total of posts	171	549
Mean number (posts/thread)	M=3.0 SD=2.4	M=8.4 SD=5.0
Mean number (posts/student)	M=4.2 SD=1.9	M=13.7 SD=3.1

scalability, performance, and interoperability, which may bring considerable repercussions on the discussion process' performance and outcomes as their lack impedes the normal learning flow (Caballé et al., 2007g).

5.4.1 From a centralized to a distributed approach

Previously reported experience (see previous section and Caballé et al., 2007g) on the use of the DF for the support of two discussion activities resulted very successful altogether from the pedagogical point of view as it showed the benefits of providing an adequate information and knowledge management in supporting the discussion process. Indeed, the quantity and quality of the contributions during the debate greatly increased in comparison to the experiences achieved using the well-known asynchronous threaded discussion forum offered by the virtual campus of the UOC (see Table 5.8). First experience was run in a centralized fashion by a single node. On the other hand, second experience was run in a distributed fashion. Next, both experiences are reported separately.

5.4.1.1 Centralized approach

The first experience using the DF was supported by using just one physical node (Windows 2003 server, Intel Pentium 3 CPU 800 MHz 512MB RAM) to support all the work on server side, namely the apache server supporting the PHP code, all the web-services and the database. Many inconveniences arose due to the overuse of the Windows server node by not only the participants of this experience but also many other students who carried out their activities misusing and abusing this server as an academic resource. As a result, the debate was interrupted several times due to node's failures. Moreover, the debate's participants suffered from serious lack of performance due to both the concurrency of different participants trying to gain access to the DF at the same time and the server's resource consumption by external users. As a result, this generated a lot of frustration and complains about not being able to make progress on the discussion process (see participants' comments in the questionnaire in Table 5.9). Finally, the lack of integration of the DF within the virtual campus forced students to authenticate again when entering the DF.

Table 5.9 shows the results of a structured and qualitative report conducted at the end of the first experience addressed to the DF' users who were also asked to compare it to the standard well-known tool they had already used in previous debates.

Table 5.9: Excerpt of a questionnaire's results on the first experience using the Discussion Forum tool supported by just one server.

Selected questions	**Average of structured responses (0 - 5)**	**Excerpt of students' comments**
Assess the DF as a collaborative tool	2	”Apart from serious technical problems, the DF fulfilled my expectations”
Evaluate how the DF fostered your active participation	3	”The system performed very slowly, I don't understand why the university is not able to provide us with a more powerful server!”
Did the DF help you acquire knowledge on the debate's issue?	4	”The standard tool is a chaos for large debates (...) DF encouraged me to participate”
Compare the DF to the campus' standard discussion tool	3	”The DF is a powerful tool but most of times I couldn't even enter because of timeout problems”

5.4.1.2 Distributed approach

The second experience was supported by the distributed version of the DF. The distribution procedure was the following (see Caballé et al., 2007g and Figure 5.13):

The DF prototype was supported by three nodes located in two separated buildings of the UOC. Each node had very different configurations:

- Linux Red Hat 3.4.6-3 cluster, Intel Xeon CPU 3.00 GHz 4GB RAM
- Windows 2003 server, Intel Pentium 3 CPU 800 MHz 512MB RAM
- Linux SuSE 2.4.21-99 machine, Intel Pentium 4 CPU 2.00 GHz, 256MB RAM

For the purpose of this experience, all Web-services of the DF prototype were replicated on each node (see subsection 4.2 in Chapter 4 for an overview of how to use the CLPL to take advantage of the distributed technology). Moreover, the same client code in the form of PHP running on Apache Web servers was installed in two nodes (Windows server and Linux SuSE machine). Finally, in this prototype, just a single instance of the database was installed in Windows server. The Windows server acted also as an entry proxy by redirecting at HTTP level all the requests received to either itself or the Linux Red Hat cluster. In this first version of the approach the database is supported by just one node, which makes the system fully dependent from it. In future iterations of this approach, it is planned to distribute the database in several nodes and manage its consistency by the data web-services. The ultimate goal in this initial version was to prove the feasibility of the distributed approach.

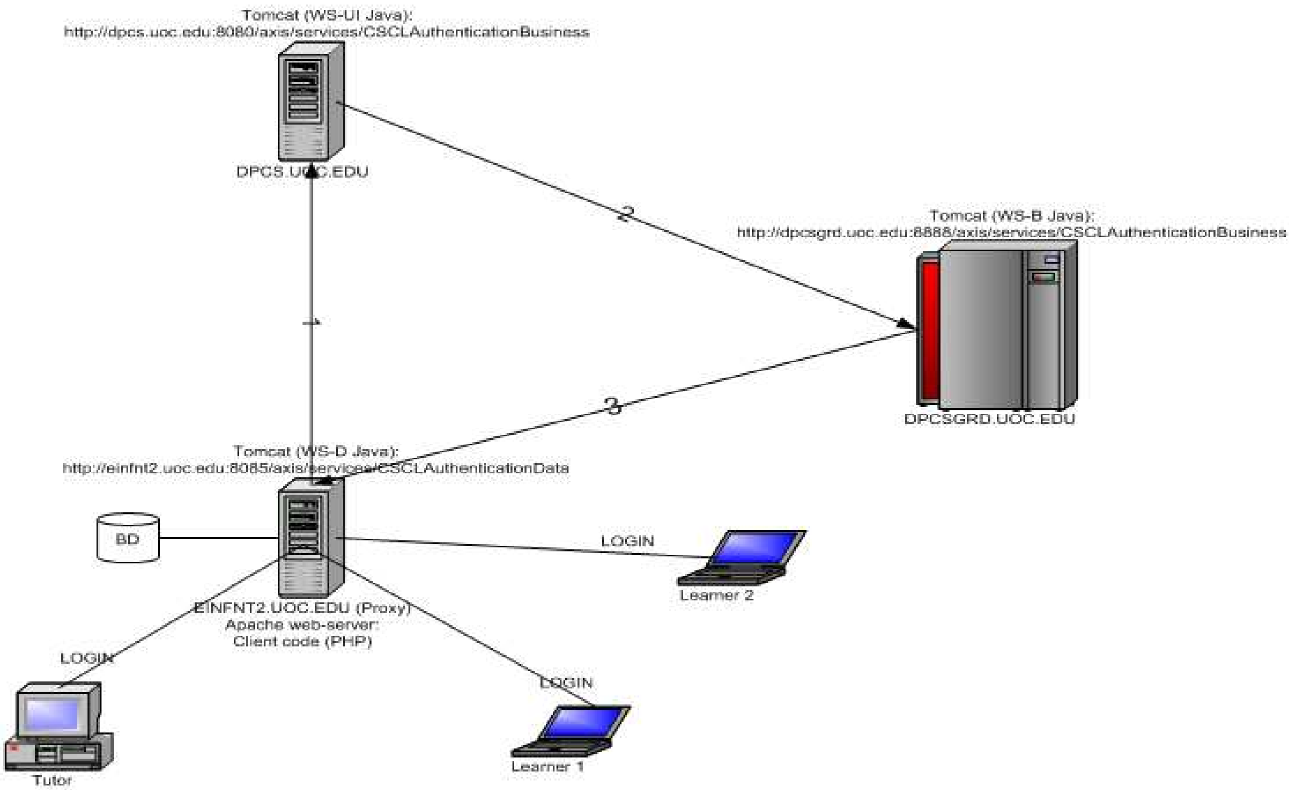

Figure 5.13: The distribution of the DF using the CLPL.

To that end (see Figure 5.13), upon the reception of a user's request, the Windows server proxy first pings at Linux SuSE machine whether it is alive. If so, the Linux SuSE machine starts dealing with the request by executing its PHP code, otherwise the Windows server itself is doing so by executing its own PHP code.

From the client PHP code it actually starts the sequential call chain of web-services for each layer, namely the user interface, business, and data web-services for each function requested. Each call to a web-service implies, if possible, to send the current request to another node. However, before calling a web-service on a different node a ping is sent to check the node's availability. If the other two possible nodes are down, the node managing the current web-service calls the next web-service locally and tries again to find another node where to call the web-service of the next layer.

When the request gets finally the data layer (i.e., the data web-service), the call is addressed from any node to the Windows server. Once the information has been successfully managed in the database, the response is sent back to the client through the same way the request took.

5.4.1.3 Experimental results

Despite the functionality provided was the same as the previous experience, the results improved from both the participants' and tutor' standpoint (Caballé et al., 2007g). Indeed, the system performed smoothly and just one time the DF was reported to be unavailable. This improvement came mainly from the utilization of other nodes apart from the Windows server, which was still overused. This fact provided an important performance gain that all students appreciated a great deal (see Table 5.10). This improvement influenced on the discussion process in terms of more contributions of better quality in average (see Table 5.11).

On the other hand, the lack of integration of the DF into the virtual campus of the UOC forced users to authenticate again and gain access to the DF. Despite not being a major inconvenience, satellite applications may impact very negatively on information systems in terms of redundancy and lack of consistency (Caballé et al., 2007g). However, the inherent interoperability feature of the CLPL's Web-services leads to a potential solution to that problem for the specific case of the DF, which is currently to do with administrative incompatibilities more than a technical issue.

Table 5.10 shows the results of the structured and qualitative report conducted at the end of the second experience. This report was the same as that conducted at the end of the previous experience.

Table 5.11 shows a comparative study between the first and second experience. Certain key indicators, such as the tutor assessment and the participation impact, improved showing the effect of the distribution approach in the learning process. Particularly interesting is the improvement of the passivity indicator showing the contributions in average pending to read. The reason may be found on the normalization of the performance of the system, which allowed the participants to spend time reading others' contributions. This, in turn, enhanced the discussion process by increasing the cogniscitive level of the discussion.

These experimental results should be taken carefully as more validation process needs to be undertaken. Nevertheless, the results here presented leads to believe that the use of the CLPL platform for enhancing the effectiveness of complex

Table 5.10: Excerpt of a questionnaire's results on the second experience using the distributed Discussion Forum tool.

Selected questions	**Average of structured responses (0 - 5)**	**Excerpt of students' comments**
Assess the DF as a collaborative tool	4	”The system performed much better I could realize its potential”
Evaluate how the DF fostered your active participation	5	”Finally the technical problems seem to have been solved and I could participate at my pace”
Did the DF help you acquire knowledge on the debate's issue?	5	”The statistical data and quality assessment displayed influenced my participation”
Compare the DF to the campus' standard discussion tool	4	”There is still more work to do to improve the user interface but the system performs well”

Table 5.11: Main learning indicators extracted from both experiences.

Indicators	**First experience**	**Second experience**
Tutor assessment 0-10 (average)	6.2	7.8
Peer assessment 0-10 (average)	5.4	6.5
Participation impact (average)	+1.8	+4.1
Passivity (pending to read on average)	88.3%	31.9%

Table 5.12: Main statistics results from the class assignment using both discussion tools.

Statistics	**Standard tool**	**CoLPE**
Number of students	43	43
Number of threads	29	17
Total of posts	174	93
Mean number (posts/thread)	M=6.0 SD=2.7	M=5.5 SD=4.5
Mean number (posts/student)	M=4.0 SD=1.6	M=2.2 SD=3.8

collaborative learning processes becomes a reality. In particular, the results shows the suitability of this platform in taking great advantage of distributed infrastructure to overcome important barriers arisen during the learning process in the form of non-functional requirements.

5.4.2 Grid infrastructure to meet non-functional requirements

This section reports again on an experience using CoLPE that took place last academic term at the UOC. As described previously 43 participants were directly involved form the course Methodology and Management of Computer Science Projects (see previous section and Caballé, at al, 2008e). For the purpose of this new perspective of the experience, other 43 participants are added forming the audience of another discussion on the same topic held at the same time in a different virtual classroom of the same course and using the standard discussion tool existing in the classroom. By comparing both discussion dynamics ruled by the same conditions sheds more light on the effects of the centralized approach in the learning experience. Next, it is presented the experimental setting and data gathered using CoLPE to support a discussion process and above all its effects in the learning experience that motivated this study.

The whole experience was supported by a Zope server (Latteier et al., 2007) on the back end, which run on a single node (i.e., Linux SuSE 2.4.21-99 machine, Intel Pentium 4 CPU 2.00 GHz, 256MB RAM) hosted at the UOC as part of its internal network.

Despite the experience was quite successful in terms of the CoLPE's knowledge-based strategy, which impacted positively on the discussion process (see section about CoLPE), the statistical results comparing this tool to the standard forum of the UOC showed that the discussion using CoLPE was poorly participative (see Table 5.12). Moreover, the results (see Table 5.13) of a structured and qualitative report conducted at the end of the discussion confirmed the undermined effect of CoLPE's centralized approach in the learning experience.

In particular, the problems were originated as follows. First, Zope is a powerful server that demands a fairly amount of hardware resources to run. Second, the need to process and analyze large and complex information collected from

Table 5.13: Excerpt of a questionnaire's results on CoLPE's evaluation to support the discussion process.

Selected questions	Average of structured responses (0 - 5)	Excerpt of students' comments
Asses CoLPE as a collaborative tool	3	”CoLPE shows great potential but I avoid to use it since it performed very badly and my time is very limited. I can't afford to spend one hour just to send one contribution and read others' messages!”
Evaluate how CoLPE fostered your active participation	3	”The standard tool is a chaos for large debates (...). Apart from many technical problems, CoLPE encouraged me to participate”
Did CoLPE help you acquire knowledge on the debate's issue?	2	”CoLPE is a powerful tool but most of times I couldn't even accede because of timeout problems”
Compare CoLPE to the campus' standard discussion tool	1	”The system performed very slowly, I don't understand why the university is not able to provide us with a more powerful server!”

users' interaction and present the knowledge extracted (see Figure 5.14) in (almost) real time caused CoLPE to perform very poorly. Third, during the rush hours, the growing number of users who concurrently requested CoLPE's knowledge-related data-intensive functionalities generated noticeable performance repercussions on the underlying hardware supporting the system. Finally, the server was down once for a few hours during the rush time due to maintenance of the internal network.

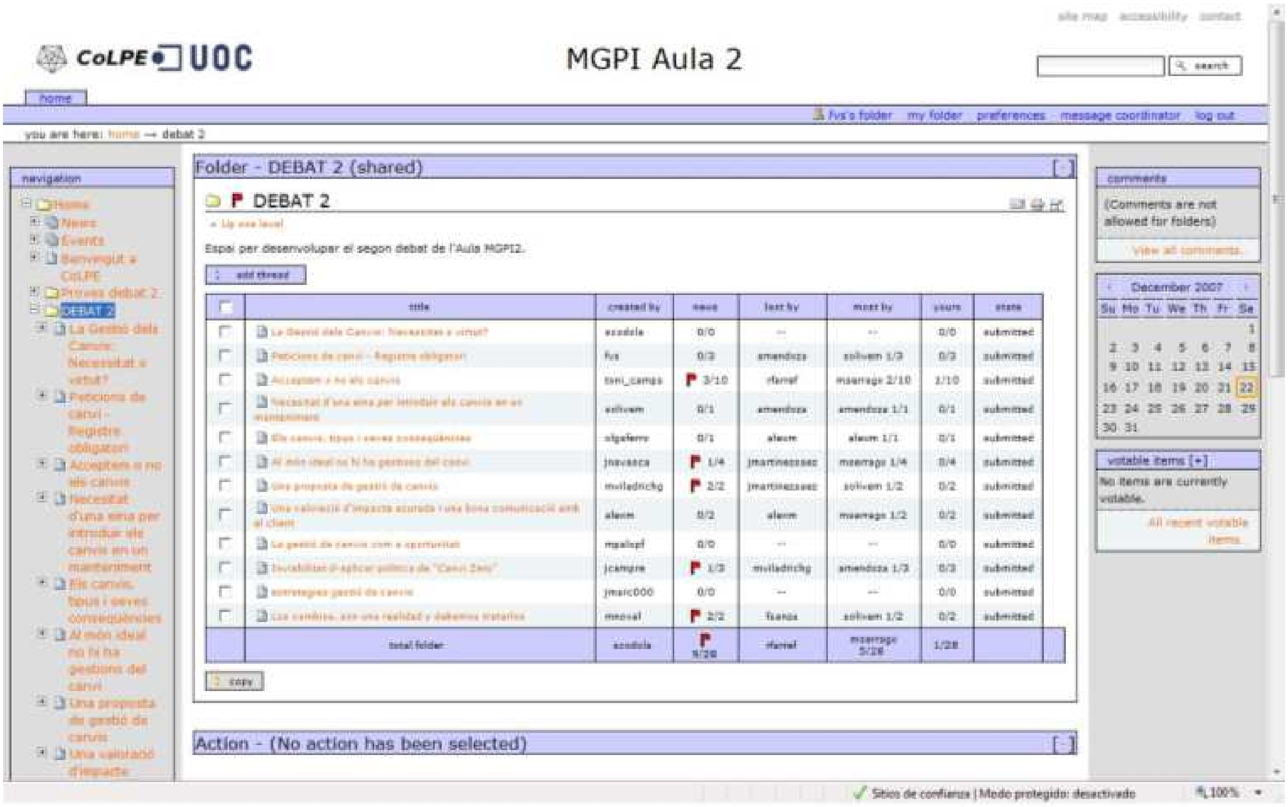

Figure 5.14: Partial feedback presented to all participants.

As a consequence of this centralized approach, important non-functional requirements could not be completely satisfied in terms of fault-tolerance, scalability, interoperability and performance. The negative impact on the discussion process caused by the lack of fault-tolerance, user scalability and interoperability is addressed in subsection 4.2.1 while the performance repercussions caused by the large amount of complex information about group activity to be processed is addresses in the next subsection 4.3. The latter is certainly an essential and extensive issue to be addressed so as to improve the overall system's performance.

Indeed, the information stored in very large log files and databases is often found with a certain degree of redundancy, tedious and ill-formatted as well as incomplete as at some cases certain user actions do not generate any log entry (e.g. user may leave CoLPE by either closing or readdressing the browser) and have to be inferred. As a consequence, treating this information is very costly in terms of time and space needing a great processing effort.

In order to meet this requirement and based on the conceptual approach investigated in Chapter 4 (see section 4.3) a Grid-based parallel approach is proposed next to specifically process log files efficiently with the aim of providing relevant and selected knowledge about group activity to be presented to participants for awareness, feedback, monitoring and so on. This issue represents one of the cornerstones of this thesis and for this reason it is addressed extensively in the next section. See section 4.3 in Chapter 4 for an extensive overview of the architectural and paradigmatic issues involved. Next section takes a technological view to completely realize the conceptual approach.

5.5 Using Grid infrastructure to parallelize the processing of log files

Based on the investigations and advances carried out in section 4.3 in Chapter 4, this section presents the implementation of the architecture built to parallelize the processing of log files by means of Grid technology. First, some preliminaries are considered and then different Grid implementations are applied to validate the approach. Finally, a discussion on the main experimental results achieved terminates this section.

5.5.1 Preliminaries

The experiences reported above using the DF and CoLPE tools shown the need for powerful infrastructure to support an efficient interaction data analysis and management and provide the discussion process and its participants with relevant knowledge about the collaboration. However, due to the exploratory nature of these experiences and the relative small number of participants, the amount of interaction data collected in log files was not large. Even though it was reported the system's poor performance caused, among other reasons, by the data-intensive functionalities of this tools, the small size of log files (e.g., the largest daily log file generated was of 1MB in size) did not justified the use of a powerful infrastructure, such as Grid. Indeed, for too small size of log files, the overhead introduced by the Grid's transmission protocol when sending the file parts to be processed in parallel is noticeable and the implemented list scheduling strategy may be spending too much time waiting for completion (see a complete report on this issue in Paniagua, Xhafa, Caballé, and Daradoumis (2005)).

Hence, in order to validate the Grid-based approach proposed here, much larger log files are to be used instead. In particular, the UOC's and BSCW's log files collect the information from thousands of users who constantly interact with the system resulting in large amounts and variety of events enlarging the size of log files (e.g., a typical UOC's daily log file may be up to 1.8 GB in size (Caballé et al., 2007a).

On the other hand, the parallelization strategy presented in section 4.3 of Chapter 4 and realized here using a Grid approach is generic and can be applied to parallelize the structuring of collaborative application's events log data (Paniagua et al., 2005).

For these reasons, in this study, log files of the UOC and BSCW will be used to validate the gain in performance while parallelizing the processing of log files from the DF and CoLPE tools by using Grid infrastructure.

5.5.2 Grid infrastructure

This section shows how the MW paradigm is appropriate for processing log files of group activity in a Grid environment (see section 4.3 in Chapter 4 and Caballé et al., 2008a for a complete overview). Since different degrees of granularity available are possible, and, moreover, there is no need for synchronization between the worker processors as tasks are completely independent from one another.

To this end, a minimal Grid implementation prototype was written using both the standard Globus Toolkit (GT)[13] middleware and an ad hoc middleware called Juxta-CAT [14] and have deployed it on the PlanetLab[15] platform. The latter is first described next.

PlanetLab is an open platform for developing, deploying and accessing planetary-scale services. At the time of this writing, PlanetLab is composed up of 852 nodes hosted in 428 different sites. Each Planetlab node runs the same base software, basically a modified Linux operating system offering services to create virtual isolated partitions in the node, which look to users as the real machine. The next subsection introduces two different realizations based on this architecture in the form of Grid middleware to efficiently parallelize the processing of logs files.

5.5.2.1 Using standard Grid middleware

The Globus Toolkit (GT) (Sotomayor and Childers, 2006; Paniagua et al., 2005) is the actual de facto Grid middleware standard. Newest version is GT4. Version 3 of GT (GT3) is a refactoring of version 2 in which every functionality is exposed to the world via a Grid service (i.e. basically, stateful web services). The core of the GT is a Grid service container implemented in Java that leverages and extends the Apache's AXIS web services container.

In order to test this Grid prototype log files of the BSCW system were used due to their relatively small size and relatively low occurrence of complex events but with high variability of file size, which fits well in this case. To this end, Planetlab is turned into a Grid fabric by installing the GT3's Grid service container. Moreover, the worker was implemented as a simple Grid service that was deployed on the GT3's container. Then, a simple Java client was developed that plays the role of the master by dispatching tasks just by calling the operations exposed by the worker Grid services, as follows:

- The *worker* Grid service publishes an interface with only one operation that the master calls in order to dispatch a task to the worker. This operation, which is implemented by wrapping the Java code of the mentioned *EventExtractor* routine (see subsection 4.3 in Chapter 4 and Appendix B), passes as an input a textual representation of the events to be processed by that task and returns a data structure containing performance information about the task executed (i.e. elapsed time, number of events processed and number of bytes processed).

- The *master* is just a simple Java application that reads from a configuration file (1) the folder that contains the event

[13]Globus: http://www.globus.org (web page as of April 2008).
[14]Juxta-CAT: https://juxtacat.dev.java.net/ (web page as of April 2008).
[15]PlanetLab: http://www.planet-lab.org (web page as of April 2008).

log files to process, (2) the available workers, (3) the number of workers to use, and (4) the size of the task to be dispatched to each worker expressed in number of events. The master then proceeds as follows: it picks as much workers as needed from the configuration file and puts them all in a queue of idle workers. Then it enters a loop reading the events from the event log files and, each time it has read a number of events, it either waits for a worker if the queue is empty or calls the worker's operation. Once the call to the worker returns, the worker is put back into the queue of idle workers. The master exits the loop when all events in the event log files have been read and all the tasks that were dispatched have finalized.

As stated, this is not a real GT3 Grid implementation of the MW paradigm but a proof-of-concept prototype, thus important features in a real environment such as fault-tolerance and dynamic discovery of available workers, are missing.

The experimental results of processing log files using GT middleware are shown later on in the subsection of experimental results.

5.5.2.2 Using ad hoc JXTA-based Grid middleware

This subsection introduces briefly the main aspects of the Grid platform, called Juxta-CAT (Esteve and Xhafa, 2006), which was used for the processing of log files. The Juxta-CAT platform has been developed using the JXTA[16] protocols (Xhafa et al., 2007) and offers a shared Grid where client peers can submit their tasks in the form of java programs stored on signed jar files and are remotely solved on the nodes of the platform. Juxta-CAT Project and its official web site have been hosted in Java.NET community[17]. In order to test this Grid prototype the very large and ill-structured log files of the UOC virtual campus were used due to the great flexibility provided by the JXTA protocols, which allowed to split the large log files into many short samples consisting of representative daily periods with different activity degrees.

The architecture of Juxta-CAT platform (Esteve and Xhafa, 2006) is made up of two types of peers: common client peers and broker peers.

- *Client peers* create and submit their requests using a GUI-based application (see Figure 5.15) and are the end users of the Juxta-CAT, which are obtained by downloading and installing the application from the official page of Juxta-CAT. Once the machine is converted into a client peer, the user will connect to the peer-to-peer network and can submit execution requests to their peer group nodes. Also, client peers will be able to process received requests sent to them by other nodes through the brokering and notify them the result of the requests, once they are completed.

- *Broker peers* are the administrators of the Grid, which are in charge of efficiently assigning client requests to the Grid nodes and notify the results to the owner's requests. Whenever a broker receives a request, it explores the state of the rest of nodes currently connected to the network, examining their working and connection statistics. Then, it uses this historical/statistical data to select, according to a price-based economic model, the best candidate peer for processing that request. To assure an efficient use of resources, brokers use an allocation algorithm, which can

[16]JXTA is found at: http://www.jxta.org/ (web page as of April 2008).
[17]Official Juxta-CAT's web site in Java.NET is found at: https://juxtacat.dev.java.net/ (web page as of April 2008).

be viewed as a price-based economic model, to determine the best candidate node to process each new received request.

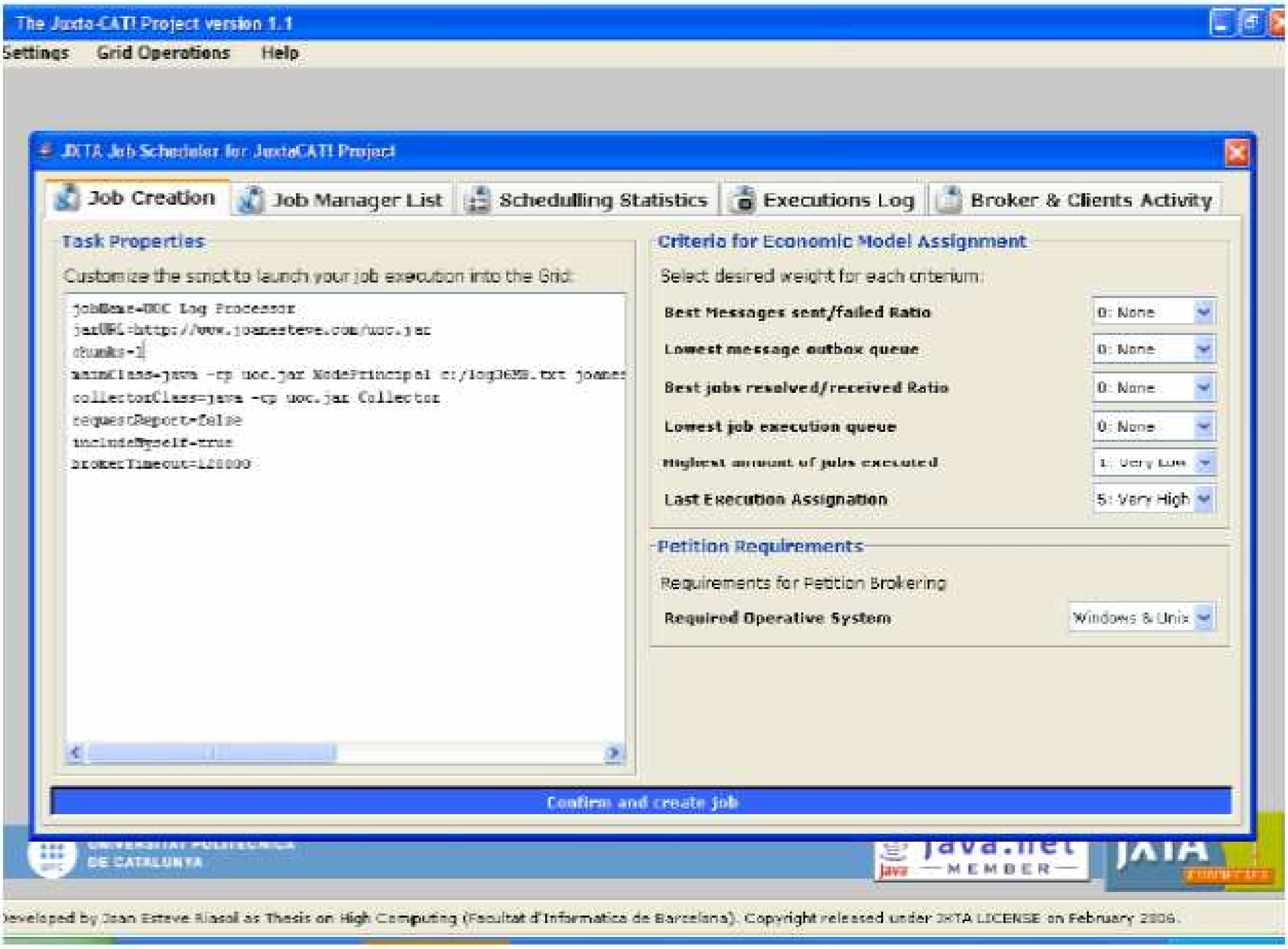

Figure 5.15: GUI view of submitting the processing of a log file to the Juxta-CAT.

The implementation and design of peers, groups, job and presence discovery, pipe-based messaging, etc. are developed using the latest updated JXTA libraries (currently release is 2.3.7) and JDK 1.5 version.

The discussion turns now to how the processing of log files is done in the Juxta-CAT platform. The implementation follows the well-known Master-Worker paradigm. Please note first that the sequential java class of the *UOCLogsProcessing* routine to process UOC's log files described in subsection 4.3 in Chapter 4 (see also Appendix B) also encapsulates functionalities to provide the division of the log file into as many equal parts as Grid nodes will be used for processing them; these parts will be later on submitted for processing to the Juxta-CAT. The main steps that would follow the user (the master node) to process a log file in the Juxta-CAT are as shown in Figure 5.16:

The *UOCLogsProcessing* routine is compiled in a unique java jar packages, which includes the library developed by Jakarta Apache needed for the FTP transfer. The code was optimized using Java ProGuard 3.5[18] so that the final jar file size is 28.7 KB. Figures 5.15 and 5.17) show the submission of a request to Juxta-CAT and the state information once

[18]ProGuard SourceForge project is found at: http://sourceforge.net/projects/proguard/ (Web page as of April 2008)

1. **[Preparation phase]**: Provide the necessary information (to the Master) for the preparation of the petitions to submit to the Juxta-CAT:
    a. Indicate the path to the log file and its name and the number of nodes participating in the processing. Log processor routines count the total number of lines of the log file, `totalNbLines`, and knowing the number of Grid nodes to be used, `nbNodes`, each node will read and process a `totalNbLines/nbNodes` of lines from the file.
    b. Indicate an FTP server, a user name and a password as well as a public address where the parts of the file will be uploaded. The implementation of FTP for Java, known as PureFTP, is included in the Jakarta Apache *commons-net-1.4.1.jar* library.
2. **[Master Loop]**: Repeat
    a. Read `totalNbLines/nbNodes` lines
    b. Upload the file to the indicated public address via FTP
    c. Create a petition and submit it to Juxta-CAT

    Until the original log file has been completely scanned.
3. **[Juxta-cat processing]**:
    a. Each time a petition is received by brokers of Juxta-CAT, it is assigned to a peer node of the platform.
    b. The peer node, upon receiving the petition, reads according to the petition's description, the part of the file it has to read via HTTP. The peer runs *UOCLogProcessing* functionality for processing the lines of the file, one at a time, and stores the results of the processing in a buffer.
    c. The peer node, once the processing of the petition is done, sends back to the master node the content of the buffer.
4. **[Master's final phase]**: Receive messages from peers and append the new received resulting file to the final file containing the information extracted from the original log file.

Figure 5.16: Algorithm to process a log file in Juxta-CAT.

it is processed. Note that the user has to just provide the in formation needed in Step 1 (see Figure 5.16); the rest is automatically done by Juxta-CAT. The experimental results using Juxta-CAT are shown later on in the next subsection.

Figure 5.17: State information of a request once it is processed.

5.5.3 Experimental results

In order to carry out a comparative study between the sequential and Grid approaches, a specific test battery was designed in which both large amounts of event information and well-stratified short samples were used. This subsection presents the experimental results achieved of the Grid prototypes while next section analyzes certain important aspects of these results to be considered.

5.5.3.1 Results from parallelizing the processing of BCSW log files

In order to test the GT-based prototype and compare the results to the sequential approach, the *EventExtractor* routine (see above in the subsection related to the sequential approach) for processing log files from the BSCW system on this Grid platform in the Planetlab nodes. To this end, as mentioned previously, it was used existing daily log files making up the whole group activity generated during a whole academic term in the course ”Software Development Techniques” at the UOC. Other tests involved a few log files with selected file size and event complexity forming a sample of each representative stratum.

The linearity found in processing time in the sequential approach (see Figure 4.2) simplified greatly the experiment by using the same event log file as input for all the Grid tests in the experiment. Then, the parameters regarding both the

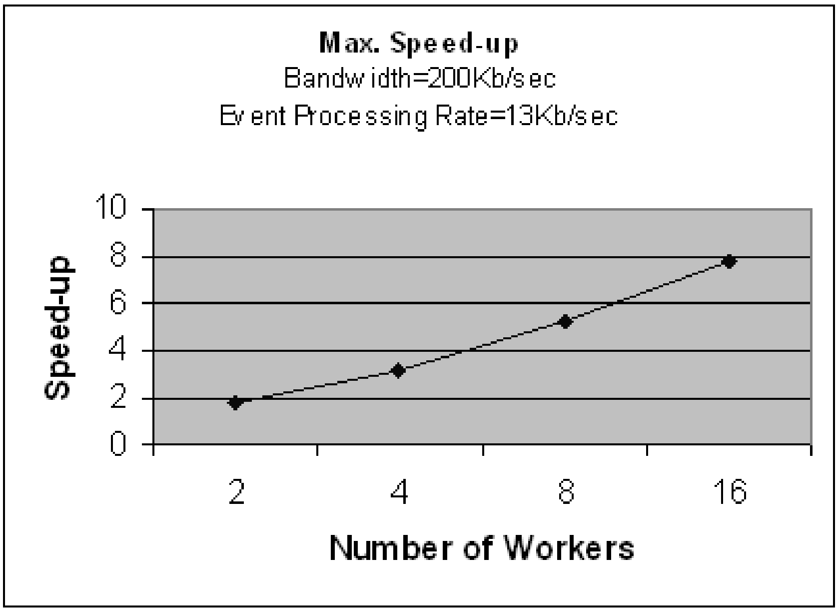

Figure 5.18: Maximum speed-up vs. number of workers.

number of workers and the size of the tasks (expressed in number of events) were left to vary. These parameters were then to be executed by the workers. Tests were run for a different number of workers with different task sizes.

Parallel speed-up is used to measure the performance gain from a parallelized execution over its serial execution and defined as

`S(s,p) = Ts (s) / Tp(s,p)`,

where *s* is the size of the log file, *Ts(s)* is the total running time of the sequential execution for a log file of size *s* and *Tp(s,p)* is the total running time of the parallel execution for a log file of size *s*, using *p* processors.

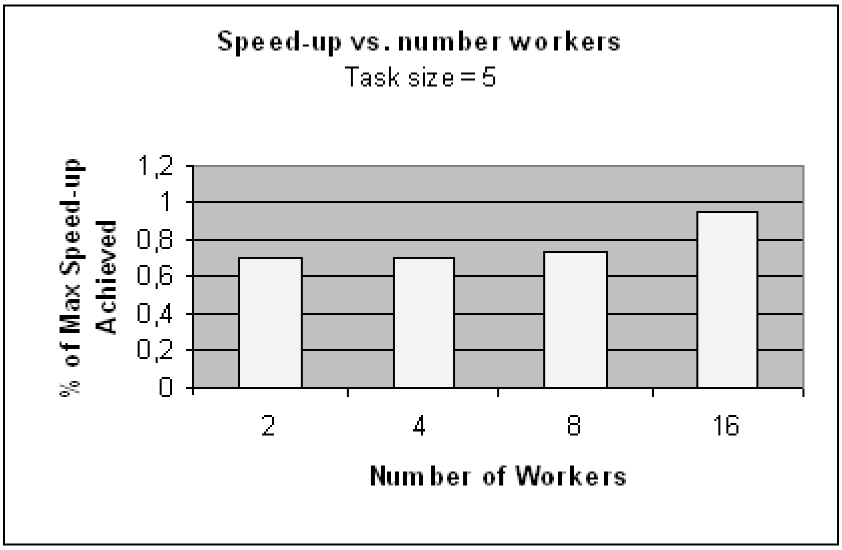

Figure 5.19: Relative speed-up vs. number of workers for a task size of 5 events.

Figure 5.18 shows the maximum speed-ups achieved for the observed bandwidth between the master processor and

Table 5.14: PlanetLab nodes used to run the experiment.)

Host	Description
planet1.manchester.ac.uk	University of Manchester
lsirextpc01.epfl.ch	École Fédérale de Lausanne
planetlab1.polito.it	Politecnico di Torino
planetlab1.info.ucl.ac.be	University of Louivain
planetlab2.upc.es Universitat	Politècnica de Catalunya
planetlab1.sics.se	Swedish Institute of Computer Sci.
planetlab1.ifi.uio.no	University of Oslo
planetlab3.upc.es Universitat	Politècnica de Catalunya
planetlab1.ls.fi.upm.es	Universidad Politécnica de Madrid
planetlab1.hiit.fi	Technology Institute of Helsinki
planetlab-1.cs.ucy.ac.cy	University of Cyprus
planetlab1.ru.is	University of Reykjavik
planetlab2.sics.se	Swedish Institute of Computer Sci.
planetlab1.mini.pw.edu.pl	Telekomunikacja Polska Warsaw
planetlab1.cs.uit.no	University of Tromso
planetlab-02.ipv6.lip6.fr	Laboratoire d'Informatique de Paris

the Planetlab nodes at the time of running the experiment and for the different number of workers tested.

As mentioned before, the worker returns the elapsed time of its execution, whereas the master executes all the events found up to the input event log files have been completely parsed and all dispatched tasks have been completed. The observed speed-up was computed for the test by dividing (1) the sum of all the elapsed times returned by each invocation of the worker into (2) the elapsed time the master run multiplied by a normalization factor to compensate the different speed between the machine running the master and the Planetlab nodes running the workers.

Therefore, the main experimental results from the parallel processing of log files are given in terms of how much close each set of workers is to achieve its theoretic maximum speed-up (see Figure 5.18) for different task size processed and, thus, providing the best processing time possible while parallelizing the data processing. To this end, Figure 5.19 shows the graphical representation of an extract of these results in relative terms for a sample of a specific 5-event size task.

5.5.3.2 Results from parallelizing the processing of UOC log files

This subsection presents the experimental results obtained after running the JXTA-based Grid platform Juxta-CAT in the Planetlab nodes on a test battery made up of the log files from the UOC virtual campus showing the speedup achieved.

This test battery uses both large amounts of log information (i.e. daily log files) and well-stratified short samples consisting of representative daily periods with different activity degrees (e.g. from 7 p.m. to 1 a.m. as the most active lecturing period and from 1 a.m. to 7 a.m. as the period with least activity in the campus). In addition, other tests involved a few log files with selected file size forming a sample of each representative stratum. This managed to provide reliable statistical results using an input data size easy to use.

The battery test was processed by the *UOCLogsProcessing* application described peviously and executed several times first on single-processor machines involving usual configurations and with different workload in order to have more reliable results in statistical terms involving file size, number of log entries processed and execution time along with other

Table 5.15: Parallel speed-up and efficiency.

Log file size	Speed-up	Efficiency
12 MB	6.1	38.2 %
24 MB	7.4	46,2 %
36 MB	9.1	56.8 %

basic statistics. Then, the same battery test was processed in a parallel fashion by Juxta-CAT using different number of nodes, specifically, 2, 4, 8, and 16 nodes, using the following nodes of the PlanetLab nodes (see Table 5.14).

Parallel efficiency measures the degree of utilization of the computing resources involved in the parallel computation and is defined as the parallel speed up divided by the number of computing resources (i.e. processors):

`E(s,p) = S(s) / p.`

From the execution times presented in Figure 5.20 and the formulas previously introduced, Table 5.15 shows the gain in terms of parallel speed-up and efficiency achieved.

5.5.4 Analysis of the results

Analyzing the experimental results obtained from the Grid prototypes, it was found that, on the one hand, from certain values of the task size, the speed-up observed was very close to the theoretic maximum achievable. Therefore, only for a very small value of the task size the impact on the speed-up can be great due to the cost of the transmission overhead. However, it was also observed that the more workers were used in the tests the closer to the theoretic maximum was the speed-up achieved by the small tasks, and this increased quickly up to the point that, given a sufficient number of workers, even the smallest tasks (i.e. one-event task size) achieved considerable speed-up.

On the other hand, the homogeneous behavior observed in Planetlab nodes justified the decision of testing with the same task size for all workers. However, in a real Grid environment, task sizes should be adjusted per worker node case to fit the dynamically changing workloads the nodes may be experimenting and to account for different machine speeds.

It was noted, however, that although the results of this experiment are promising, a deeper and more precise analysis on both the primary interaction occurring between participants in the virtual classrooms and the real collaborative learning activity based on complex parameters of the collaboration, such as the above-mentioned task performance, group functioning and scaffolding, it is expected to generate a much larger amount and more complex events than those used for the experiments . This scenario will take much more advantage of the benefits provided by a Grid environment and will provide a more useful knowledge about the actual performance occurring in the on-line learning activity and will help monitor and support learning participants more conveniently.

5.5.5 Final discussion

After this study, it is important to keep in mind what the motivation of this whole study is. To this end, this last section summarizes the main results achieved and points out very briefly how to apply them in a learning context.

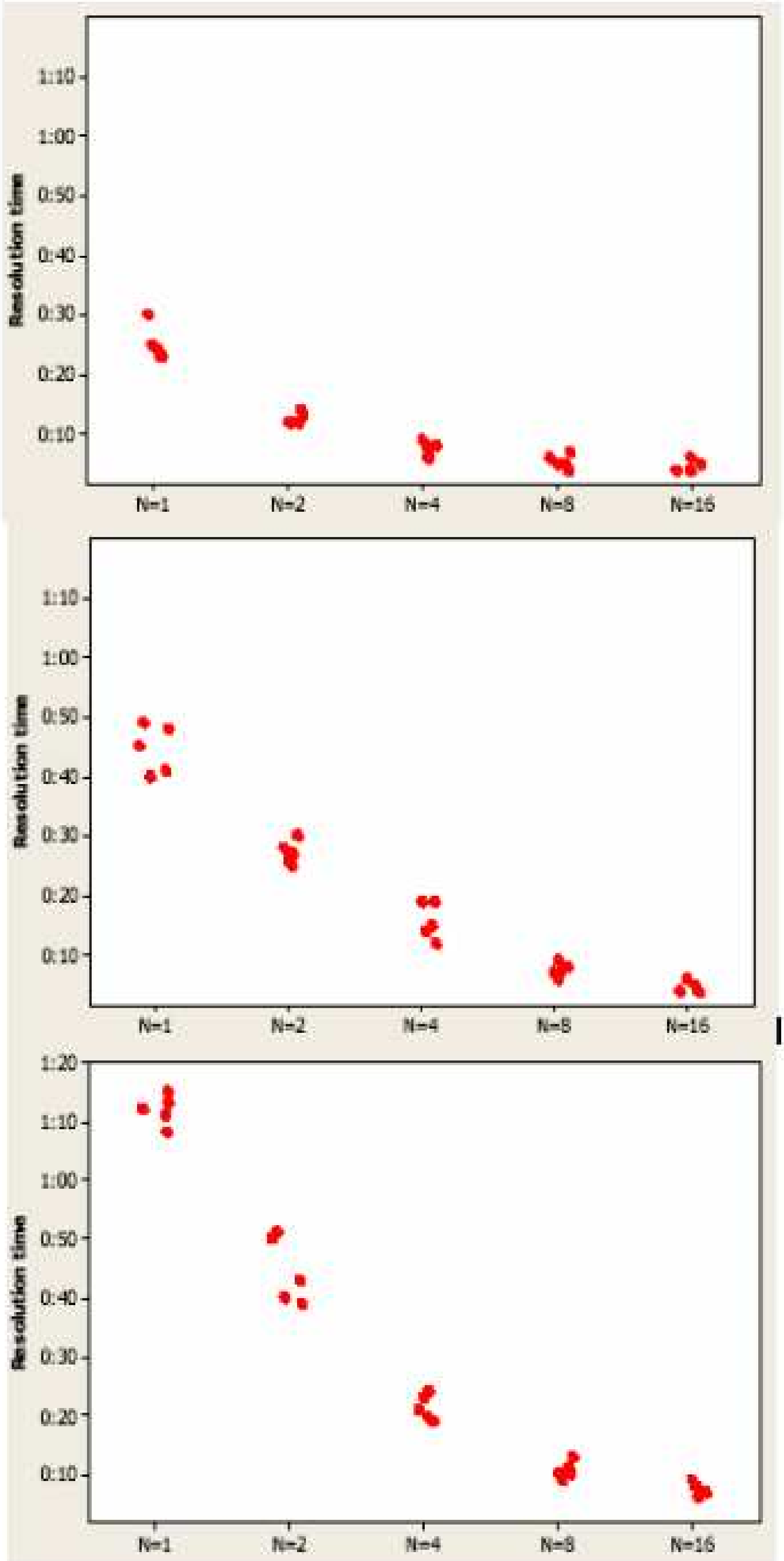

Figure 5.20: Three execution time results for log files with sizes of 12MB, 24MB and 36MB respectively; x-axis indicates the number of processors and y-axis the processing time (mm:ss).

Figure 5.21: An example of awareness (in the form of flags) and complex feedback (statistics information) constantly presented to students from the interaction information collected, processed and analyzed during a discussion process.

So far, it has been argued how the provision of continuous knowledge to on-line teams in CSCL environments can greatly improve the group activity in terms of decision-making, group organization, social engagement, support, monitoring and so on. As a result, large amounts of log information generated from the collaborative interaction need to be efficiently processed and analyzed. Moreover, in order to make the knowledge extracted from the analysis be useful for awareness and feedback purposes (see Figure 5.21), users should be provided with both single information delivered fast in (almost) real-time and complex, exhaustive, yet structured deferred information thus stressing even more the processing requirements beyond those of a single computer.

Furthermore, Web-based applications that support on-line distance learning may be also greatly benefited from the extraction of selected knowledge from the log files for user modeling purposes (Caballé et al., 2007c). Indeed, these applications, due to the high degree of user interaction, take great advantage of the tracking-based techniques of user modeling such as providing broader and better support for the users of Web-based educational systems (Gaudioso, 2003). The data analysis of the information captured from the actions performed by learners is a core function for the modeling of the learner's behavior during the learning process and of the learning process itself as well. In addition, the building of learner models may help identify navigation patterns (see Figure 5.22) and adapt the system's usability to the actual learners' needs resulting in a great stimulation of the learning experience (Caballé et al., 2007d).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1								PERIOD (03/18/2006 1:00:15 - 03/19/2006 1:00:00)								
2		~24H (1:00:15 - 1:00:00)			~6H (1:00:15 - 6:59:59)			6H (7:00:00 - 12:59:59)			6H (13:00:00 - 18:59:59)			6H (19:00:00 - 1:00:00)		
3		Abs	%		Abs	%		Abs	%		Abs	%		Abs	%	
4	SESSIONS															
5	Total sessions	39848	100		1809	100		11216	100		14305	100		12518	100	
6	Sessions with visits	31403	78		1368	75		8893	79		11254	78		9888	78	
7	Empty visits	8445	21		441	24		2323	20		3051	21		2630	21	
8	Session time accumulated	6915H 18M 59S			298H 0M 53S			1879H 40M 26S			2450H 30M 4S			2287H 7M 36S		
9	Session time mean	0H 10M 20S			0H 10M 3S			0H 10M 11S			0H 10M 24S			0H 11M 5S		
10	Total logins	39848	100		1809	100		11216	100		14305	100		12518	100	
11	Explicit logouts (by button Exit)	12307	30		594	32		3623	32		4312	30		3778	30	
12	Implicit logouts (readdress/close browser)	27541	69		1215	67		7593	67		9993	69		8740	69	
13	Logins campus CAT	29211	73		1099	60		8288	73		10693	74		9131	72	
14	Logins campus IB	8797	22		609	33		2436	21		2984	20		2768	22	
15	Logins others	1840	4		101	5		492	4		628	4		619	4	
16	Total	39848	100		1809	100		11216	100		14305	100		12518	100	
17	VISITS (out of total visits)			MEAN TIME/VISIT			MEAN TIME/VISIT			MEAN TIME/VISIT			MEAN TIME/VISIT			MEAN TIME/VISIT
18	Classrooms	29809	44	0H 1M 56S	1223	42	0H 2M 0S	8305	44	0H 1M 52S	10815	45	0H 1M 55S	9466	45	0H 2M 1S
19	Mailbox	14264	21	0H 2M 25S	689	23	0H 2M 27S	4204	22	0H 2M 8S	5038	21	0H 2M 15S	4333	20	0H 2M 36S
20	Start	15509	23	0H 4M 49S	674	23	0H 4M 35S	4350	23	0H 4M 41S	5661	23	0H 4M 44S	4824	23	0H 4M 55S
21	Tutorship	421	0	0H 3M 10S	24	0	0H 2M 49S	106	0	0H 3M 46S	147	0	0H 3M 11S	144	0	0H 2M 54S
22	Community areas	682	1	0H 3M 40S	44	1	0H 3M 37S	203	1	0H 4M 6S	251	1	0H 3M 5S	184	0	0H 3M 26S
23	Library	741	1	0H 3M 30S	36	1	0H 3M 33S	204	1	0H 2M 48S	270	1	0H 3M 40S	231	1	0H 3M 43S
24	Agenda	2323	3	0H 3M 15S	93	3	0H 4M 38S	671	3	0H 2M 56S	844	3	0H 3M 25S	715	3	0H 2M 59S
25	Help	34	0	0H 4M 21S	2	0	0H 6M 55S	9	0	0H 5M 40S	13	0	0H 1M 0S	10	0	0H 3M 30S
26	News	290	0	0H 3M 24S	11	0	0H 6M 1S	95	0	0H 3M 52S	102	0	0H 3M 17S	82	0	0H 4M 7S
27	Directory	1255	1	0H 3M 42S	66	2	0H 4M 44S	304	1	0H 3M 9S	443	1	0H 3M 21S	442	2	0H 3M 13S
28	Personalization	273	0	0H 4M 24S	6	0	0H 5M 10S	65	0	0H 5M 39S	112	0	0H 3M 1S	90	0	0H 2M 54S
29	Secretary's office	837	1	0H 4M 17S	38	1	0H 2M 18S	254	1	0H 3M 50S	292	1	0H 4M 32S	253	1	0H 3M 50S
30	Total	66438	100		2906	100		18770	100		23988	100		20774	100	
31	FIRST VISIT (out of total sessions with visit)															
32	Classrooms	18267	58		750	54		5059	56		6606	58		5852	59	
33	Mailbox	9315	29		454	33		2804	31		3281	29		2776	28	
34	Start	1711	5		84	6		450	4		602	5		575	5	
35	Tutorship	103	0		3	0		25	0		36	0		39	0	
36	Community areas	190	0		11	0		63	0		70	0		46	0	
37	Library	329	1		9	0		83	0		124	1		113	1	
38	Agenda	761	2		25	1		217	2		274	2		245	2	
39	Help	12	0		2	0		3	0		3	0		4	0	
40	News	41	0		1	0		18	0		14	0		8	0	
41	Directory	384	1		18	1		85	0		146	1		135	1	
42	Personalization	61	0		0	0		16	0		17	0		28	0	
43	Secretary's office	229	0		11	0		70	0		81	0		67	0	
44	Total	31403	100		1368	100		8893	100		11254	100		9888	100	
45	PATTERNS (max 1 pattern/session. Over total sessions)															
46	Classrooms+Mailbox	2456	6		139	7		682	6		837	5		798	6	
47	Mailbox+Classrooms	6429	16		295	16		1796	16		2306	16		2032	16	
48	Classrooms+Start+Mailbox	421	1		15	0		121	1		161	1		124	0	
49	Mailbox+Start+Classrooms	1464	3		74	4		450	4		544	3		396	3	
50	PROCESSING EFFORT															
51	File size (bytes)	1011281675			47127976			272732240			358410732			333010727		
52	Log lines processed	1970084			91890			572481			722929			582784		
53	Processing time (sec)	1423,21			63,7			403,74			522,12			465,33		
54																

Figure 5.22: Analysis results of a 24-hour UOC campus' log containing about 40,000 user sessions. The results show a variety of navigation information and patterns for user modeling purposes.

The several Grid approaches reported in this section represent an important step toward overcoming demanding requirements by improving the processing time of a large amount of log files storing complex event information from the group activity. The results obtained and the experience achieved in these studies lead to conclude that the question whether the Grid is beneficial or not will heavily depend on the volume and structure of information being processed. Therefore, these results encourage to keep up working on the development of a real working Grid implementation to address the problem of processing group activity event log files. The ideas and experiences gathered so far will provide an essential background and resources to face forthcoming challenges in this thesis.

5.6 Effective and timely development of CSCL systems

This section reports on the experience achieved in using the Collaborative Learning Purpose Library (CLPL) systematically for the construction of CSCL applications. To this end, several undergraduate students participated in developing small applications by extensively reusing the CLPL in all stages of the development. Results show an increase in production and quality of the applications while guiding the whole software engineering process.

First subsection offers an overview of the CLPL from the development technology and reuse standpoints (see Chapter 3 for a complete description). Next subsection reports the experience at the Open University of Catalonia (UOC) of using the CLPL in the development of final projects in Computer Science degree.

5.6.1 Advanced software engineering using the CLPL

The merge of software engineering paradigms and techniques described in Chapter 3 (i.e., Service-Oriented Architecture (SOA), Model-Driven Development (MDD), Generic Programming (GP) and Product Line Architectures (PLA), yielded the development of the CLPL, which provides developers with the latest engineering software techniques for the effective and timely development of CSCL applications (Caballé, 2008d).

The CLPL is based on the GP and Service-Oriented Architecture (SOA) and Generic Programming (GP) paradigms so as to enable a complete and effective reutilization of its generic components as a skeleton for the construction of any collaborative learning application by means of implementing the conceptualization of the fundamental needs existing in any collaborative learning experience (Caballé et al., 2007e). The CLPL also provides full support to distribution, reusability, flexibility and interoperability as key aspects to address the current non-functional needs in software development in general, and specifically in the CSCL domain (Caballé, 2007d).

There a great deal of similarities between the pervasive and challenging collaborative learning needs and the benefits provided by SOA (Caballé et al., 2008d). As a result of this matching, SOA appears to be the best choice to support the development of the CLPL. Indeed, SOA enhances educational organizations by increasing the flexibility of their pedagogical strategies, which can be continuously adapted, adjusted, and personalized to each specific target learning group. Moreover, SOA facilitates the reutilization of successful collaborative learning experiences and makes it possible for the collaborative learning participants to easily adapt and integrate their current best practices and existing well-known learning tools into new learning goals. Over the last years, CSCL has become a complex and extensive domain. Therefore, the application of the GP principles appear to be a good choice for the development of the CLPL by, first, identifying those parts which are common to most applications of the CSCL domain. Then, proceed to isolate the fundamental parts in the form of abstractions from which the basic requirements are obtained. Finally, encourage the greatest possible reusability of the resulting generic components for the construction of as many CSCL applications as possible.

In order to turn the CLPL into an effective software platform, its development was based on the Model-Driven Development(MDD) approach. This paradigm fits very well in combination with the GP and SOA principles due to the clear separation of a generic, reusable technology-independent model from a different, flexible technology-dependent implementation models. Moreover, Web-services are the implementation technology chosen for the CLPL given the widely adopted protocols and standards, which represents the very rationale of the this technology. These entire standards represent a suitable context to guarantee interoperability and scalability by taking great advantage of the distributed technologies.

Finally, in order to automate as much as possible the transition from the Platform Independent Model (PIM) to the appropriate Platform Specific Model (PSM) (see Chapter 2 for an overview on PIM and PSM), the latest research results in this thesis are leading to deal with XMI files (see OMG, 2004 for details), which are XML-tagged files as the result of coding UML diagrams. In combination with XSL style sheets, it is possible to turn the PIM's XMI files into WSDL files, which represent the input for a Web-service working environment to transform them into a specific-language architecture design (PSM). Lack of comply with standard of the existing UML case tools is the major problem to face in the future as well as how to provide a more complete and detailed realization of the desired PSM. These open questions are to be investigated as further work in the context of this thesis.

5.6.2 Experience and results in developing applications systematically with the CLPL

Final project (i.e. thesis) courses offered by the UOC dispose specific areas related to engineering software and in particular the development of software applications for collaborative learning. The interested area is called Web-based Applications for Collaborative Work and it is intended to provide the needed resources and framework in support the students who develop collaborative tools for e-learning. The main support is two-fold, the lecturer's guidance during the whole development and the organization of the course's currila in a few deliverables that students are required to submit in deadline fashion. These deliverables are planned to fit the different phases of the traditional software development process (i.e., specification, design, and implementation) plus both an initial stage to plan and organize the whole project and a thesis' defense at the end of the course. It is worth mentioning here that final project courses present a high dropout in comparison to other courses[19], which influenced this experience a great deal.

Three undergraduate students of the UOC have chosen so far to develop their thesis in this area using the CLPL. Despite all of them dropped out because of personal reasons, they could perform part of work. Representative efforts were two applications, namely a collaborative agenda and calendar intended to ease the personal and group planning as

[19]Because of the particular profile of the UOC (students are about 30 years old on average and 95% with a job), the dropout ratio is about 50%.

well as a document repository to help organize the document-based outcomes generated by the group activity.

Each of these students was recommended to use and reuse the CLPL as much as possible from the very first step of the development. As a result, the identification of the requirements and their analysis by reusing the UML diagrams of the CLPL were highly satisfactory and of good quality. Students reported saving time and effort by avoiding to start from scratch but having 75% on average of the development already fulfilled instead. Most importantly, they reported to feel highly confident in developing the applications since the CLPL provided them with strong guidance and support in terms of going through the different stages of the software development and the UML modeling at any stage. The deliverables submitted [20] were of the high quality, submitted in time and assessed accordingly.

Despite promising, these results are not conclusive due to the exploratory nature of the approach. More experiences are expected to come and validate the CLPL as the *de facto* platform to support students in this final project area for the timely and effective development of CSCL tools of high quality (Caballé et al., 2008d).

[20] An example of a deliverable is found at: http://clpl.uoc.edu/docs/AgendaSpecifications.pdf (Web page as of April 2008). Please note the student's personal data have been removed not to disclose the anonymate.

Chapter 6

Conclusions and future work

This thesis describes a multi-dimensional approach for enhancing knowledge management and scaffolding in a collaborative learning environment that contributes to the improvement of the collaborative learning process.

To this end, first, a conceptual model is described that captures and classifies the main types of information generated in group activity by means of potential indicators of effective collaboration while taking a global group well-being function into account. Then, a process for interaction data management guides in how this information can be transformed into useful knowledge to be presented to the interested parties. This entire conceptual framework is finally translated into a computational model of collaborative learning interaction resulting in the main contribution of this thesis in form of a generic platform called Collaborative Learning Purpose Library (CLPL) that provides the skeleton and necessary resources for the systematic, robust and effective construction of CSCL applications from all development stages. In particular, by means of specific knowledge management support this platform can be used to embed relevant and selected information into CSCL applications in an efficient manner. The CLPL is designed and realized in a way to take great advantage of distributed computing, and in particular Grid infrastructure.

The CLPL library has been validated by means of the successful construction of several CSCL applications in support for the discussion process occurring in the virtual classrooms of the Open University of Catalonia. The timely and effective construction of these and other CSCL applications lead to prove the feasibility of this platform in terms of advanced software engineering techniques. Distributed infrastructure has been then added to validate the suitability of the CLPL in distributed environments and satisfactorily meet demanding non-functional requirements appearing in modern virtual campuses, and in particular in the experiences run in the context of this thesis. One essential requirement for the success of the online collaborative learning activity is the efficient embedding of large amounts of complex information and knowledge about the ongoing group activity into collaborative learning environments. In this thesis, an extensive conceptual and empirical Grid-aware approach have been shown for processing log files of group activity in an efficient yet simple manner.

The merge of these synergies represents an attractive but quite laborious challenge that in the context of this thesis have yielded CSCL systems capable of enhancing knowledge management that contributes to the improvement of the discussion process in virtual collaborative learning environments. To this end, several experiences of innovative knowledge-based structured discussion forums have been reported. Despite the results of these experiences are not conclusive due to its exploratory nature, from the analysis of the results it has been proved to promise significant benefits for students in the context of learning by discussion in higher education, and in education in general.

Therefore, the experimental results lead to conclude that the entire multi-approach achieves the goals and hypothesis formulated in this thesis and especially encourage to keep working and exploring further this direction. Next, it is described how the main problems have been addressed and what solutions have been proposed.

6.1 Main thesis' achievements

The main objective in this thesis is the exploration of the importance of an efficient management of information generated from group activity in CSCL practices for its further use in extracting and providing knowledge on interaction behavior. To this end, the first step was to investigate a conceptual model for data analysis and management so as to identify the many kinds of indicators that describe collaboration and learning and classify them into high-level potential categories of effective collaboration. Indeed, it was found more evident key discourse elements and aspects than those shown by the literature (see Chapter 1), which played an important role both for promoting student participation and enhancing group and individual performance, such as, the impact and effectiveness of students' contributions, among others, that were explored in this work (see Chapter 2). By making these elements explicit, the discussion model proposed accomplished high students' participation rates and contribution quality in a more natural and effective way (see Chapter 5). This approach went beyond a mere interaction analysis of asynchronous discussion in the sense that it built a multi-functional model that fostered knowledge sharing and construction, developed a strong sense of community among students, provided the tutor with a powerful tool for students' monitoring, discussion regulation, while it allowed peer facilitation through self, peer and group awareness and assessment.

The results of the research described so far motivated the development of a computational system as the translation from the conceptual model into a computer system the implements the management of the information and knowledge acquired from the group activity, so as to be efficiently fed back to the collaboration. The achievement of a generic, robust, flexible, interoperable, reusable computational model that meets the fundamental functional needs shared by any collaborative learning experience was largely investigated in this thesis (see Chapter 3). The systematic reuse of this computational model permitted a fast adaptation to new learning and teaching requirements, such as learning by discussion, by relying on the most advanced software engineering processes and methodologies from the field of software reuse, and thus important benefits were obtained in terms of productivity, quality, and cost.

Therefore, another important objective in this thesis was achieved by means of exploring and extending suitable software reuse techniques, such as Generic Programming, so as to allow the computational model to be successfully particularized in as many as situations as possible without losing efficiency in the process. In particular, based on domain analysis techniques, a high-level computational description and formalization of the CSCL domain were identified and modeled. Then, different specific-platform developments that realized the conceptual description were provided. It was also explored a certain level of automation by means of generation-based techniques based on Service-Oriented Architectures and Web-services while passing from the conceptual specification to the desired realization, which greatly facilitated the development of CSCL applications using this computational model.

Based on the outcomes of this exploration and experimental results (see Chapter 3 and 5, and Appendix A), the computational systems built (see Appendix B for a quick overview) were capable of managing both qualitative and quantitative information and transforming it into useful knowledge for all the implicated parties in an efficient and clear way. This was achieved by both the specific assessment of each contribution by the tutor who supervised the discussion and by rich statistical information about student's participation. This statistical data was automatically provided by the system; for instance, statistical data shed light on the students' engagement in the discussion forum or how much interest drew the student's intervention in the form of participation impact, level of passivity, proactivity, reactivity, and so on. The aim was to provide both a deeper understanding of the actual discussion process and a more objective assessment of individual and group activity (see Chapter 5).

This information was then processed and analyzed by means of a multivariate statistical model in order to extract useful knowledge about the collaboration. The knowledge acquired was communicated back to the members of the learning group and their tutor in appropriate formats, thus providing valuable awareness and feedback of group interaction and performance as well as may help identify and assess the real skills and intentions of participants. The most important result from the conceptual model for interaction data analysis and management was a great improvement and enhancement of the learning and teaching collaborative experiences (see Chapter 5 and Appendix A).

Finally, in order to achieve another important objective, the possibilities of using distributed and Grid technology to support real CSCL environments were also extensively explored in this thesis (see Chapter 4). The results of this investigation led to conclude that the features provided by these technologies form an ideal context for supporting and meeting demanding requirements of collaborative learning applications. This approach was taken one step further for enhancing the possibilities of the computational model in the CSCL domain and it was successfully adopted on an empirical and application basis (see Chapter 4). From the results achieved (see Chapter 5), the use of distributed technologies considerably enhanced and improved the collaborative learning experience. In particular, the use of Grid computing was successfully applied for the specific purpose of increasing the efficiency of processing a large amount of information from group activity log files (see Chapter 4 and 5).

6.2 Further research

Future work will focus first on investigating how to incorporate information retrieval and document filtering techniques into the stage of information collection to automatically extract knowledge from information with a high degree of informality. Secondly, how to make an in-depth analysis through data mining techniques so as to provide tutors with ongoing progress of students learning during the discussion activity. In overall, it is planned to investigate how to integrate a portable, general and reusable CSCL ontology into the CLPL platform as a declarative representation of the knowledge embedded into CSCL applications with the aim to formally model and describe how these applications are built and hence understand better how real learning groups work.

Next steps will be to investigate the state of the art and study the feasibility of incorporating Natural Language Processing (NLP) techniques (Tennant, 1982) as part of the inference engine of the process of managing interaction data about group activity (see subsection 2.3 in Chapter 2). The aim is to semi-automate the analysis and findings of the specific discussion participants' intentions and skills without having to explicitly categorize their contributions.

Finally, it is plan to explore the shift from traditional e-learning to mobile learning (m-learning) (Grew et al., 2007), which has come to play a major role in educational environments by taking advantage of the extensively used mobile and wearable technology (such as podcasting) to provide anywhere, anytime learning. Both theoretical frameworks and best real practices adopting ubiquity and pervasiveness will be investigated in support for formal and informal communities of learning practice, such as the Open University of Catalonia.

REFERENCES

Ali, A., Anjum, A., Azim, T., Thomas, M., Steenberg, C., Newman, H., Bunn, J., Haider, R., Rehman, W. (2004). *JClarens: A Java Based Interactive Physics Analysis Environment for Data Intensive Applications*. In proceed. of ICWS 2004. pp. 716-723

Aloisio G., Cafaro M., Epicoco I., Fiore S., Lezzi D., Mirto, M., and Mocavero, S. (2005). *Resource and Service Discovery in the iGrid Information Service*. In proceedings of ICCSA'05.

Amin, K., Nijsure, S., and von Laszevski, G. (2002). *Open Collaborative Grid Services Architecture (OCGSA).* In Proceedings of the W3C EuroWeb 2002 Conference, Oxford, UK.

Amoretti M, Reggiani M., Zanichelli F., Conte G., (2005). *Enabling Service-Oriented Grids using a Peer-to-Peer Approach*. In proceedings of WETICE'05.

Anido-Rifón, L., Llamas, M., Fernández, M. J., Caeiro, M., Santos, J., Rodríguez, J. (2001). *A Component Model for Standardized Web-based Education*. Conference Proceedings of 10th International World Wide Web Conference, Hong Kong (China), ACM Press. pp 86-95. ISBN 1-58113-348-0.

Aspeli, M. (2007). *Professional Plone development*. PACKT publishing.

Avouris, N., Komis, V., Fiotakis, F., Margaritis, M., Tselios, N. (2003). *On tools for analysis of collaborative problem solving*. Proceedings of the 3rd IEEE International Confer-ence on Advanced Learning Technologies.

Ateveh, K. and Lockemann, P. C. (2006). *Reuse- and Aspect-Oriented Courseware Development*. Educational Technology and Society, 9 (4), 95-113. Bacelo Blois, A., Becker, K. (2002). *A Component-Based Architechture to Support Collaborative Application Design.* In Haake and J. Pino Eds., Groupware: Design, Implementation and Use. LNCS, Vol. 2440, pp. 134-143.

Baloian, N., Berges, A., Buschmann, S., Gaßner, K., Hardings, J., Hoppe, U. and Luther, W. (2002). *Document Management in a Computer-Integrated Classroom*. In Haake and J. Pino Eds., Groupware: Design, Implementation and Use. LNCS, Vol. 2440, pp.35-44.

Bentley, R., Hosrtmann, T., Trevor, J. (1997). *The World Wide Web as enabling technology for CSCW: The case of BSCW*. In: Computer-Supported Cooperative Work: Special Issue on CSCW and the Web, vol. 6, pp. 111-134.

Bell, J., Bellegarde, F., Hook, J., Kieburtz, R. B., Kotov, A., Lewis, J., McKinney, L., Oliva, D. P., Sheard, T., Tong, L.,Walton, L.and Zhou, T. (1994). *Software Design for Reliability and Reuse: A Proof-of-Concept Demonstration*. In

Proceedings of the Conference TRI-Ada.

Bote Lorenzo, M.L., Gómez Sánchez, E., Vega Gorgojo, G., Dimitriadis, Y., Asensio Pérez, J.I., Jorrín Abellán, I.M. (2007) *Gridcole: a tailorable grid service based system that supports scripted collaborative learning.* Computers and Education. Academic Press: Elsevier Ltd, May 2007.

Baloian, N., Galdames, P., Collazos, C., and Guerrero, L. (2004). *A model for a Collaborative Recommender System for Multimedia Learning Material*. In proceedings of the 10th International Workshop on Groupware. Berlin: Springer.

Barros, M., Verdejo, M. (2000). *Analysing student interaction processes in order to improve collaboration*. The DEGREE approach. International Journal of Artificial Intelligence in Education, 11, 221-241.

Bentley, R., Horstmann, T., and Trevor, J. (1997). *The World Wide Web as enabling technology for CSCW: The case of BSCW*. Computer-Supported Cooperative Work: The Journal of Collaborative Computing, 2-3. Amsterdam: Kluwer Academic Press.

Bratitsis, T. and Dimitracopoulou, A. (2006). *Monitoring and Analysing Group Interactions in asynchronous discussions with the DIAS system*. Lecture Notes in Computer Science, LNCS 4154. (pp. 54-61). Heidelberg-Berlin: Springer-Verlag. ISBN: 978-3-540-39591-1.

Bushey, R., Mauney, JM., and Deelman, T. (1999). *The Development of Behaviour-Based User Models for a Computer System.* In Judy Kay (ed.), User Modeling. In Proceed. of the UM'99. Springer Wien New York, pp. 109-118.

Caballé S. and Xhafa F. (2003a). *A Study into the Feasibility of Generic Programming in Java*. Proceedings of the Second International Conference on the Principles and Practice of Programming in Java. pp. 25-26. Kilkenny, Ireland. ACM Press. ISBN: 0-9544145-1-9.

Caballé, S., and Xhafa, F. (2003b). *A Study into the Feasibility of Generic Programming for the Construction of Complex Software*. Proceedings of the 5th Generative Programming and Component Engineering, Erfurt, Germany. ISBN: 3-9808628-2-8.

Caballé S., and Xhafa, F. (2003c). *Disseny i implementació d'una llibreria de components de propòsit genèric per als usuaris d'un sistema informàtic. (in Catalan)*. Bachelor's thesis in Computer Science, Open University of Catalonia. Thesis awarded with honors.

Caballé, S., Xhafa, F., Daradoumis, T., and Marquès, J.M. (2004). *Towards a Generic Platform for Developing CSCL Applications Using Grid Infrastructure*. Proceedings of the First International Workshop on Collaborative Learning Applications of Grid Technology, Chicago, IL, USA.

Caballé, S., Daradoumis, T., Paniagua, C., and Xhafa, F. (2005a). *A Grid Approach to Provide Effective Awareness to On-line Collaborative Learning Teams*. Proceedings of the First International Kaleidoscope Learning GRID Special Interest Group Workshop on Distributed e-Learning Environments. Naples, Italy.

Caballé, S., Paniagua, C., Xhafa, F., and Daradoumis, Th. (2005b). *A Grid-aware Implementation for Providing Effective Feedback to On-line Learning Groups*. Proceedings of the Second International Workshop on Grid Computing and its Application to Data Analysis, Agia Napa, Cyprus. Lecture Notes in Computer Science, Vol. 3762, p. 274-284.

Caballé, S., Xhafa, F. (2005c). *A Web-Based Structured Discussion Forum for Collaborative Learning Environments*. Master's thesis in Computer Science, Open University of Catalonia. Thesis awarded with honors.

Caballé S., Daradoumis Th. (2005d). *Survey on Service Oriented Frameworks for e-Learning*. Kaleidoscope WP 12 Learning Grid SIG. Deliverable D12.1.6. pp. 4-7.

Caballé, S., Daradoumis, Th., Xhafa, F.(2006). *A Model for the Efficient Representation and Management of Online Collaborative Learning Interactions*. In P. Cunningham and M. Cunningham (Eds.), Building the Knowledge Economy: Issues, Applications and Case Studies. pp. 1485-1492. IOS Press, Amsterdam, Netherlands. ISBN: 1-58603-682-3.

Caballé, S., Daradoumis, Th., Xhafa, F., Esteve J. (2007a). *Supporting Effective and Useful Web-based Distance Learning*. In proceedings of the Third International Conference on Web Information Systems and Technologies (WEBIST 2007), Barcelona, Spain. INSTICC Press. ISBN: 978-972-8865-79-5. pp. 536-539.

Caballé, S. (2007b). *On the Advantages of Using Web and Grid Services for the Development of Collaborative Learning Management Systems*. Proceed. of the First International Workshop on P2P, Parallel, Grid and Internet Computing (3PGIC-2007). Vienna, Austria. IEEE Computer Society. ISBN: 0-7695-2823-6.

Caballé, S., Daradoumis, Th., Xhafa, F., Esteve J., Barolli, L., Durresi, A., (2007c). *Using a Grid Platform for Enabling Real Time User Modeling in On-line Campus*. Proceedings of the First International Conference on Complex, Intelligent and Software Intensive Systems (CISIS 2007). Vienna, Austria. IEEE Computer Society. ISBN: 0-7695-2823-6.

Caballé, S., Xhafa, F., Fernández, R., Daradoumis, Th., (2007d). *Efficient Enabling of Real Time User Modeling in On-line Campus*. Proceedings of the User Modeling 2007. Corfu, Greece. Lecture Notes in Computer Science, vol. 4511, Springer 2007, ISBN 978-3-540-73077-4.

Caballé, S., Daradoumis, T., and Xhafa, F. (2007e). *A Generic Platform for the Systematic Construction of Knowledge-based Collaborative Learning Applications*. Book: Architecture Solutions for e-Learning Systems. Idea Group Inc (IGI Global). Chp. XII (pp. 219-242). Press, Hershey, PA: Idea Group, ISBN: 978-1-59904-633-4.

Caballé, S., Daradoumis, Th., Xhafa, F. (2007f). *Efficient Embedding of Information and Knowledge into CSCL Applications*. Proceedings of the Second International Conference of E-Learning and Games (Edutainment 2007). Hong Kong, China. Lecture Notes in Computer Science, vol. 4469, Springer 2007, ISBN 978-3-540-73010-1.

Caballé, S.,Xhafa, F., Daradoumis, Th. (2007g). *A Service-oriented Platform for the Enhancement and Effectiveness of the Collaborative Learning Process in Distributed Environments*. Proceedings of the International Conference on Grid computing, high-performAnce and Distributed Applications (GADA 2007). Vilamoura, Algarve, Portugal. Lecture Notes in Computer Science, vol. 4804. ISBN: 978-3-540-76835-7.

Caballé, S., Xhafa, F., and Daradoumis, T. (2008a). *A Grid Approach to Efficiently Embed Information and Knowledge about Group Activity into Collaborative Learning Applications*. Book: The Learning Grid Handbook. pp. 173-197. Amsterdam, The Netherlands: IOS Press. ISBN: 978-1-58603-829-8.

Caballé, S., Daradoumis, Th., Xhafa, F. (2008b). *Providing an Effective Structured and Context-aware Asynchronous*

Discussion Forum for Collaborative Knowledge Building. In proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications (ED-MEDIA 2008). Vienna, Austria. AACE Press. Presented. To appear.

Caballé, S., Feldman, J., Thaw, D. (2008c). *Supporting Communities of Learning Practice by the Effective Embedding of Information and Knowledge into Group Activity*. In proceedings of the International Workshop on Collaborative E-Learning Systems and Applications (CESA 2008). Barcelona, Spain. IEEE Computer Society. ISBN: 0-7695-3109-1/08.

Caballé, S. (2008d). *Combining Generic Programming and Service-Oriented Architectures for the Effective and Timely Development of Complex e-Learning Systems*. In proceedings of the Second International Conference on Complex, Intelligent and Software Intensive Systems (CISIS 2008). Barcelona, Spain. IEEE Computer Society. ISBN: 0-7695-3109-1/08. Best paper award of CISIS 2008.

Caballé, S.,Xhafa, F., Daradoumis, Th., Abraham, A. (2008e). *Effective Interaction Analysis for an Effective Provision of Knowledge about the Discussion Process to CSCL Practices.* Proceedings of the International Workshop on Collaborative E-Learning Using Computational Grid (CelGrid 2008). Ostrava, The Czech Republic. IEEE Computer Society. To appear.

Caballé, S., Feldman, J. (2008f). *CoLPE: Communities of Learning Practice Environment*. In proceedings of the Directions and Implications of Advanced Computing; Conference on Online Deliberation (DIAC-2008/OD-2008). Berkeley, CA, USA. To appear.

Caballé, S., Feldman, J., Thaw, D. (2008g). *CoLPE: Support for Communities of Learning Practice by the Effective Embedding of Information and Knowledge about Group Activity*. Book: Intelligent Collaborative e-Learning Systems and Applications. Series: ”Studies in Computational Intelligence”. Berlin, Germany: Springer-Verlag. To appear.

Caballé, S., Daradoumis, Th., Xhafa, F. (2008h). *Fostering Collaborative Knowledge Building by the Effective Provision of Knowledge about the Discussion Process*. Special Issue on: Advances in Intelligent Information Management Systems and Applications. International Journal of Business Intelligence and Data Mining (IJBIDM). Inderscience. To appear.

Carbó, JM., Mor, E., Minguillón, J. (2005). *User Navigational Behavior in e-Learning Virtual Environments*. In proceed. of The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05), pp. 243-249.

Casanova, H. and Dongarra, J. (1998). *NetSolve: Network enabled solvers*, IEEE Computational Science and Engineering, 5(3) pp. 57-67.

Clark, H. and Schaefer, E. (1989). *Contributing to discourse*. Cognitive Science. 13(2), 259-294.

Czarnecki, K. *Overview of Generative Software Development* (2005). In J.-P. Banâtre et al. (Eds.): Unconventional Programming Paradigms (UPP) 2004, Mont Saint-Michel, France, LNCS 3566, pp. 313-328.

Czarnecki, K. and Eisenecker, UW (2000). *Generative Programming: Methods, Techniques, and Applications*. Addison-Wesley.

Conklin, E. J. (1992). *Capturing Organizational Memory*. D. Coleman (Ed.), Proceedings of GroupWare '92 (pp. 133-137). San Mateo, CA: Morgan Kaufmann.

Czajkowski, K., Ferguson, D., Foster, I., Frey, J., Graham, S., Maguire, T., Snelling, D., Tuecke, S. (2004). From Open Grid Services Infra-structure to WS-Resource Framework: Refactoringand Evolution. Available at: http://www.globus.org/wsrf/ (Web page as of April 2008)

Daradoumis, T. (1997). A*n integrated approach to modelling tutorial dialogues under an overall rhetorical structure*. Doctoral Dissertation, Politechnical University of Catalonia, Barcelona, Spain.

Daradoumis, T., Xhafa, F., and Marquès, J.M. (2003). *Exploring Interaction Behaviour and Performance of Online Collaborative Learning Teams*. Proceedings of the 9th International Workshop on Groupware, Grenoble, France. Springer, Lecture Notes in Computer Science, 2806, 126-134.

Daradoumis, T., Martínez, A. and Xhafa, F. (2004a). *An integrated approach for analyzing and assessing the performance of virtual learning groups*. Proceedings of the 10th International Workshop on Groupware. Berlin: Springer.

Daradoumis, T., Caballé, S., and Xhafa, F. (2004b). *CSCL SIG Kaleidoscope 2004 Symposium*. Lausanne (Suïssa), 7-9 October 2004. Providing Efficient Event Information Management to CSCL Applications. Not published.

Daradoumis, T., Xhafa, F., and Juan Pérez, A. (2005). *A Framework for Assessing Self Peer and Group Performance in e-Learning*. Roberts, T. S. (Eds), Self, Peer, and Group Assessment in e-Learning. Hershey, PA: Idea Group Press.

Daradoumis, T., Martínez, A., and Xhafa, F. (2006). *A Layered Framework for Evaluating Online Collaborative Learning Interactions*. International Journal of Human-Computer Studies. Special Issue on ”Theoretical and Empirical Advances in Groupware Research”, Academic Press: Elsevier Ltd.

De Smet, M., Van Keera, H. and Valckea, M. (2008). *Blending asynchronous discussion groups and peer tutoring in higher education: An exploratory study of online peer tutoring behaviour*. Computers and Education. 50(1): 207-223, Academic Press: Elsevier Ltd, January 2008.

De Wever, B., Schellens, T., Valcke, M., and Van Keer, H. (2006). *Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review*. Computers and Education. 46(1): 6-28, Academic Press: Elsevier Ltd, January 2006.

Dimitriadis, Y. A., Asensio, J. I., Toquero, J., Estébanez, L., Martín, T.A., and Martínez, A. (2002). *Towards a component based system for the CSCL domain* (in Spanish). In Proceedings of the Symposium on Informatics and Telecommunications, Sevilla, Spain.

Dillenbourg, P. (1999a). *What do you mean by ”Collaborative Learning”?* P. Dillenbourg (Ed.), Collaborative learning. Cognitive and computational approaches, 1-19. Oxford: Elsevier Science.

Dillenbourg, P. (1999b). *Collaborative Learning. Cognitive and Computational Approaches*. Elsevier Science Ltd, 1-19.

Ellis, C., Gibbs, S., and Rein, G. (1991). *Groupware: some issues and experiences*. Communications of the ACM, 34,1, 9-28.

El Saddik, A., Y., Georganas, N. and Ralf Steinmetz, R. (2001). *Web-based multimedia tools for sharing educational resources*. In Journal on Educational Resources in Computing (JERIC) Volume 1, Issue 1es. Article No. 9. ISSN:1531-

4278. ACM Press. New York, NY, USA.

Esteve J. and Xhafa F. (2006). *Juxta-CAT: A JXTA-based Platform for Distributed Computing.* In proceed of the ACM International Conference on Principles and Practice of Programming in Java (PPPJ'06).

Fernández, M., Kadiyska, Y., Suciu, D., Morishima, A., and Tan, W. (2002) *SilkRoute: A framework for publishing relational data in XML*. In ACM Transactions on Database Systems (TODS), Vol. 27, Issue 4.

Foster, I. and Kesselman, C. (1998). *The Grid: Blueprint for a Future Computing Infrastructure*. San Francisco, CA: Morgan Kaufmann. Gomaa, H. Designing Software Product Lines wtih UML: From Use Cases to Pattern Based Software Architectures. Reading, Massachusetts: Addison-Wesley, 2005.

Foster I., Kesselman C., Tuecke, S. (2001). *The Anatomy of the Grid: Enabling Scalable Virtual Organizations*. International Journal of Supercomputer Applications and High Performance Computing.

Gaudioso, E., Boticario, J.G. (2003). *Towards web-based adaptive learning communities*. Proceedings of Artificial Intelligence in Education 2003, Sydney, Australia. IOS Press.

Guangzuo, C., Fei, C., Hu, C., Shufang, Li. (2004). *OntoEdu: A Case Study of Ontology-based Education Grid System for E-Learning.* GCCCE journal.

Gomaa, H. (2004). *Designing Software Product Lines with UML: From Use Cases to Pattern-Based Software Architectures.* Addison-Wesley.

Goux, J.P., Kulkarni, S., Linderoth, J. and Yoder, M. (2000). *An enabling framework for master-worker applications on the computational grid.* In 9th IEEE International Sympo-sium on High Performance Distributed Computing (HPDC'00). IEEE Computer Society.

Grew, P., Pagani, E., Giudici, F. (2007). *From E-Learning to M-Learning: Architectures to Support University Teaching.* Book: Architecture Solutions for e-Learning Systems. Idea Group Inc (IGI Global). Chp. IV (pp. 62-79). Press, Hershey, PA: Idea Group, ISBN: 978-1-59904-633-4.

GuiLing W., YuShun L., ShengWen Y., ChunYu M., Jun Xu, Meilin S., (2005a). *Service-Oriented Grid Architecture and Middleware Technologies for Collaborative E-Learning.* IEEE SCC 2005: 67-74.

GuiLing W., YuShun L., ShengWen Y., ChunYu M., Jun Xu, Meilin S., (2005b). *Towards a Grid-Based Collaborative Platform for E-Learning.* WEBIST 2005: 510-517.

Gütl, C. (2007). *Moving Towards a Generic, Service-based Architecture for Flexible Teaching and Learning Activities.* In C. Pahl (Ed.) Architecture Solutions for E-Learning Systems. Chp. I (pp. 1-24). Hershey, PA: Idea Group (IGI Global). ISBN: 978-1-59904-633-4.

Gutwin, C., Stark, G., and Greenberg, S. (1995). *Support for Workspace Awareness in Educational Groupware.* Proceedings of the ACM Conference on Computer Supported Collaborative Learning, Bloomington, Indiana, USA.

Hew, K.F. and Cheung, W.S. (2008). *Attracting student participation in asynchronous online discussions: A case study of peer facilitation.* Computers and Education. Academic Press: Elsevier Ltd, (in press).

Inaba, A., Ikeda, M., and Mizoguchi, R. (2003). *What Learning Patterns are Effective for a Learner's Growth? -*

An ontological support for designing collaborative learning. Proceedings of the International Conference on Artificial Intelligence in Education, Sydney, Australia. (pp. 219-226).

Jonassen, D. H. (1994). *Learning with media: Restructuring the debate.* Educational Technology Research and Development, Vol. 42, No.2, pp. 31-39, 1994.

Koschmann, T. (1996). *Paradigm shifts and instructional technology*. In T. Koschmann (Ed.), CSCL: Theory and Practice of an Emerging Paradigm, Mahwah, New Jersey, Lawrence Erlbaum Associates, (1-23).

Kulesza, E., Reinalda, B. (2006). *The Bologna Process - Harmonizing Europe's Higher Education*. 2nd edition. Barbara Budrich Publishers. ISBN: 978-3938094532

Kyung-Soo, J. (2001). *A design of middleware components for the connection between XML and RDB*. Proceedings of IEEE ISIE 2001, pp. 1753-1756 Vol. 3, 1753 - 1756.

Laforcade, P., Choquet, C., Nodenot, T., Pierre-Andre, C. (2007). *Model-Driven Engineering (MDE) and Model-Driven Architecture (MDA) applied to the Modeling and Deployment of Technology Enhanced Learning (TEL), Systems: promises, challenges and issues*. In Architecture Solutions for E-Learning Systems, C. Pahl (Eds). 2007, IDEA Group Inc: Hershey, PA. 2004(2).

Latorre, A., del Rincón, D., and Arnal, J. (1996). *Bases metodológicas de la investigación educativa* (in Spanish). Barcelona: GR2. Lund, K. Human Support in CSCL: what, for whom, and by whom? (2004). J.-W. Strijbos, P. Kirscher and R. Martens (Eds.), *What we know about CSCL in Higher Education*, 167 -198. Dordrecht : Kluwer Academic Publishers.

Latteier, A., Pelletier, M., McDonough, C., Sabaini, P. 2007, *The Zope Book*. Available online at http://www.zope.org/Docum (Web page as of April 2008).

Marquès, J. M., Navarro, L. (2006). *LaCOLLA: A Middleware to Support Self-sufficient Collaborative Groups*. Computers and Artificial Intelligence 25(6).

Martin, J.R. (1992). *English Text: Systems and Structure*. Amsterdam: Benjamin Press.

Martínez, A., de la Fuente, P., and Dimitriadis, Y. (2003). *Towards an XML-based representation of collaborative interaction*. B. Watson, S. Ludvigsen, and U. Hoppe (Eds.), Proceedings of the International Conference on Computer Support for Collaborative Learning 2003, Bergen (pp. 379-384). Dordrecht: Kluwer Academic Publishers.

Mason, R., (1992). *Methodologies for Evaluating Applications of Computer Conferencing. Methodologies for evaluating applications of computer conferencing*. In A.R. Kaye (Ed.). Collaborative learning through computer conferencing: The Najaden papers. pp. 105-116. New York, NY: Stringer-Verlag.

MacDonald, J.(2003). *Assessing online collaborative learning: process and product*. In Int. J. of Computers and Education, 40, pp. 377-391.

McGrath, J.E. (1991). *Time, Interaction and Performance: A Theory of Groups*. Small Group Research, 22, 147-174.

McKay, A. (2006) *The Definitive Guide to Plone*. Apress.

Ochoa, S., Guerrero, L.A., Fuller, D. and Herrera, O. (2002). *Designing the Communications Infrastructure of Group-*

ware Systems. In Haake and J. Pino Eds., Groupware: Design, Implementation and Use. LNCS, Vol. 2440, pp.114-133.

OGSI (2003). *Open Grid Services Infrastructure, version 1.0, Proposed Recommendation.* Available at: http://www.ggf.org/ (Web page as of April 2008)

OMG. (2008). *Object Management Group: Model-Driven Architecture*. http://www.omg.com/mda (web page as of April 2008).

Paniagua, C., Xhafa, F., Caballé, S., and Daradoumis, T. (2005). *A Grid Prototype Implementation for Real Time Processing of Group Activity Log Data in Collaborative Applications*. Proceedings of the 2005 International Conference on Parallel and Distributed Processing Techniques and Applications, Las Vegas, Nevada, USA (pp. 1177-1183).

Pankatrius, V., Vossen, G. (2003). *Towards E-Learning Grids: Using Grid Computing in Electronic Learning*. Proceedings of IEEE Workshop on Knowledge Grid and Grid Intelligence, Halifax, New Scotia, Canada, (pp. 4-15).

Pena-Shaff, J. and Nicholls, C. (2004). *Analyzing student interactions and meaning construction in computer bulletin board discussions*. Computers and Education. 42(3): 243-265, Academic Press: Elsevier Ltd, April 2004.

Pérez, M., Carretero, J., García, F., Peña, J.M., and Robles, V. (2003) *MAPFS: A Flexible Infrastructure for Data-Intensive Grid Applications*, In 2003 Annual CrossGrid Project Workshop and 1st European Across Grids Conference, Santiago de Compostela, Spain

Pilkington, R.M. (1999). *Analysing Educational Discourse: The DISCOUNT Scheme*. Technical Report No. 99/2, Jan 1999, The University of Leeds, Leeds, England.

Puntambekar, S. (2006). *Analyzing collaborative interactions: Divergence, shared understanding and. construction of knowledge*. Computers and Education, 47, pp. 332-351.

Salomon, G. (1993). *No distribution without individual's cognition: a dynamic interactional view*. In G. Solomon (Ed.), Distributed cognitions (pp. 111-138). Cambridge, MA: Cambridge University Press.

Schellens, T. and Valcke, M. (2006). *Fostering knowledge construction in university students through asynchronous discussion groups*. Computers and Education. 46(4):349-370, Academic Press: Elsevier Ltd, May 2006.

Schrire, S. (2006). *Knowledge building in asynchronous discussion groups: Going beyond quantitative analysis*. Computers and Education. 46(1): 49-70, Academic Press: Elsevier Ltd, January 2006.

Schwartz, D. L. (1999). *The productive agency that drives collaborative learning*. In proceedings of Dillenbourg (Ed.), Collaborative learning: Cognitive and computational approaches. (pp. 197-219). NY: Elsevier Science/Permagon.

Self, J.A. (1994). *Dormobilea: vehicle for metacognition*. In: Chan, T.W. and Self, J.A. (Eds.), Emerging Computer Technologies in Education, (pp. 1-20). Charlottesville, VA: AACE.

Sfard, A. (1998). *On two metaphors for learning and the dangers of choosing just one*. Educational Researcher 27(2), pp. 4-13.

Shum, S.B. (2002). *CoAKTinG: Collaborative Advanced Knowledge Technologies in the Grid*. Proc. Second Workshop on Advanced Collaborative Environments, Eleventh IEEE Int. Symposium on High Performance Distributed Computing (HPDC-11), Edinburgh, Scotland.

Simonson, M., Smaldino, S., Albright, M., Zvacek, S. (2003). *Teaching and Learning at a Distance*. Prentice Hall, Upper Saddle River, NJ.

Soller, A. (2001). *Supporting Social Interaction in an Intelligent Collaborative Learning System*. Int. J. of Artificial Intelligence in Education, 12: 40-62.

Sotomayor, B., Childers, L. (2006). *Globus Toolkit 4: Programming Java Services*. Morgan Kaufmann Publishers / Elsevier.

Stahl, G. (2006). *Group Cognition: Computer Support for Building Collaborative Knowledge*. Acting with Technology Series, Cambridge, MA: MIT Press. ISBN: 978-0262195393.

Stahl, G. (2002). *Computer Support for Collaborative Learning: Foundations for a CSCL Community*. Proceedings of CSCL 2002. LEA: Boulder, CO (2002).

Strijbos, J-W., Martens, R., Prins, F., and Jochems, W. (2006). *Content analysis: What are they talking about?*. Computers and Education. 46(1): 29-48, Academic Press: Elsevier Ltd, January 2006.

Szyperski, C. (2003). *Component Software: Beyond Object-Oriented Programming*. 2nd ed. ACM Press.

Tennant, H. (1981). *Natural Language Processing.* Petrocelli Books, New York, NY.

Watson, P. (2003). *Databases and the Grid*. In Grid Computing: Making The Global Infrastructure a Reality, Berman, F. et al. (eds.) Wiley.

Webb, N. (1992). *Testing a theoretical model of student interaction and learning in small groups*. R. Hertz-Lazarowitz and N. Miller (Eds.), Interaction in Cooperative Groups: The Theoretical Anatomy of Group Learning, 102-119. Cambridge Univ. Press, NY.

Webber, S. (2005). *The success of open source.* Harvard University Press.

W3C Working Group (2004). *Web Services Architecture Document.* http://www.w3.org/TR/ws-arch/ (Web page as of April 2008).

Xhafa, F., Caballé, S., Daradoumis, Th., and Zhou, N. (2004). *A Grid-Based Approach for Processing Group Activity Log Files*. R. Meersman et al. (Eds): OTM Workshops 2004, Proceedings of the First International Workshop on Grid Computing and its Application to Data Analysis, Agia Napa, Cyprus. Lecture Notes in Computer Science, 3292, 175-186. Berlin: Springer.

Xhafa, F., Barolli, L., Fernández, R., Daradoumis T., and Caballé, S. (2007). *Jxta-Overlay: An interface for Efficient Peer Selection in P2P JXTA-based Systems*. Journal: Computer Standards and Interfaces. Elsevier. JCR 2007, Impact Factor: pending.

Xhafa, F., Barolli, L., Fernández, R., Daradoumis, T., Caballé, S., Kolici, V. (2008a). *Extending JXTA Protocols for P2P File Sharing Systems*. In proceedings of the Second International Conference on Complex, Intelligent and Software Intensive Systems (CISIS 2008). Barcelona, Spain. IEEE Computer Society. ISBN: 0-7695-3109-1/08.

Xhafa, F., Barolli, L., Fernández, R., Daradoumis T., and Caballé, S. (2008b). *Extension and evaluation of JXTA protocols for supporting reliable P2P distributed computing*. International Journal of Web Information Systems, Vol 4

(1), 2008, Emerald Pub.

Zaheer Abbas, Muhammad Umer, Mohammed Odeh, Richard McClatchey, Arshad Ali, Farooq Ahmad. (2005). *A Semantic Grid-based E-Learning Framework (SELF)*, In proceedings of CLAG + Grid.edu 2005, Cardiff, UK.

Zumbach, J., Hillers, A., and Reimann, P. (2003). *Supporting Distributed Problem-Based Learning: The Use of Feedback in Online Learning*. T. Roberts (Ed.), Online Collaborative Learning: Theory and Practice, 86-103. Hershey, PA: Idea.

Zumbach, J., Schönemann, J., and Reimann, P. (2005). *Analyzing and Supporting Collaboration in Cooperative Computer-Mediated Communication*. In T. Koschmann, D. Suthers, and T. W. Chan (Eds.), Computer Supported Collaborative Learning 2005: The Next 10 Years! (pp. 758-767). Mahwah, NJ: Lawrence Erlbaum.

APPENDIX A: Experimental results

This appendix presents the full results of six experiences carried out at the Open University of Catalonia (UOC) over the last 3 academic terms (i.e., Spring 2007, Fall 2007 and Spring 2008). Section 5.2 in Chapter 5 shows a detailed analysis result of one representative experience.

Five lecturers and about 370 graduate and undergraduate students in all participated directly in these experiences. Other 370 students were also involved indirectly. Experiences took place in 3 courses of the Computer Science degree at the UOC, namely, Methodology and Management of Computer Science Projects (MGPI), Management of Organizations and Computer Science Projects (GOPI), and Software Engineering (EP).

All experiences consisted in on-line discussion activities performed in class as part of courses' curricula. The topic of the discussion was to do with certain important issues related to the courses' contents. Students were required to participate depending on whether the discussion assignment was a requirement to pass the course. Otherwise students could participate at convenience. In all cases they were free to contribute to the discussion as much as they needed.

Statistical results are shown for each experience comparing the contributing dynamics supported by both the UOC's threaded standard tool and a prototype of a new discussion tool. Please note that quantitative data on number of posts do not include lecturers' contributions but students' only. Finally, students were asked to fill in a questionnaire at the end of each assignment to assess the collaboration tools supporting the discussions. Only in one experience (i.e., experience #1) the assessment was a requirement, the rest of them it was optional.

Experience #1 - Spring term 2007

- Course: Methodology and Management of Computer Science Projects (MGPI)
- Assignment type and duration: Double class discussion activity at different periods of the course for 5 weeks in all
- Participation type: Required
- Discussion tools: UOC's *standard tool* (ST) and *Discussion Forum* (DF) version 0.6 (prototype)
- Potential participants: 40 (ST) + 40 (DF) graduate students

Statistic results

See Table 6.1.

Table 6.1: Main statistics extracted from the first experience using both discussion tools.

Statistics	**ST**	**DF**
Number of students	40	40
Number of threads	57	65
Total of posts	171	549
Mean number (posts/thread)	M=3.0 SD=2.4	M=8.4 SD=5.0
Mean number (posts/student)	M=4.2 SD=1.9	M=13.7 SD=3.1

Questionnaire results

Participation (required): 38/40 (95%). Most relevant questionnaires' responses (Original language):

Questionnaire #1

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   La meva valoració és 4. Igual li dono una puntuació una mica alta però he de ser just amb l'esforç i amb la comparació amb l'eina anterior del Campus. En aquest sentit la millora és força bona. No obstant, això s'ha de millorar la usabilitat en properes versions.

2. Com creus que la nova eina ha incidit en el grau quantitatiu de la teva participació. (valora 0 -5 i fes els comentaris que creguis convenients)

   La meva valoració és 4. En aquets cas la valoració és totalment objectiva ja que la millora ha estat clara quant a les facilitats que m'ha proporcionat per participar-hi.

3. Com creus que la nova eina ha incidit en la qualitat de les teves aportacions. (valora 0 -5 i fes els comentaris que creguis convenients)

   La meva valoració és 2. En aquest cas la incidència ha estat baixa, no per l'eina que valoro bona, sinó perquè la qualitat de la meva participació, al meu entendre, no ha estat influenciada per aquesta.

4. En comparació als espais de fòrum i debat de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Avantatges:

   - Molta més facilitat per seguir els fils de discussió.

   - Possibilitat de crear diferents espais de discussió amb més claredat.

   - Informació addicional dels diferents debats i fils fonamental per ubicar-s'hi.

   - Funcionalitats extres que ajuden i fomenten la participació: estadístiques, valoracions, etc.

   Inconvenient:

   - Estar ubicada fora del Campus.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

- Ubicar-la dins el campus amb la mateix usuari i contrasenya. És a dir, no haver d'introduir-les de nou i amb un enllaç directe des del Campus.

- Millorar la navegació pels fils.

Questionnaire #2

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   5. A mí personalmente me ha ayudado mucho, ya que no tenía muy claros los conceptos.

2. Com creus que la nova eina ha incidit en el grau quantitatiu de la teva participació. (valora 0 -5 i fes els comentaris que creguis convenients)

   4. La verdad que me he involucrado mucho en esta iniciativa ya que he visto un canal muy abierto con la gente.

3. Com creus que la nova eina ha incidit en la qualitat de les teves aportacions. (valora 0 -5 i fes els comentaris que creguis convenients)

   5. Despues de leer los comentarios del foro, mis comentarios han sido mas enriquecedores con el tiempo.

4. En comparació als espais de fòrum i debat de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Ventajas: Al ser obligatorio, supongo que es mucho mas activo y enriquecedor. Inconvenientes: Ninguno.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Estaria bien que fuera un foro abierto y se hicieran debates de cualquier tema.

Questionnaire #3

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   TRES - Crec que es bona l'idea d'obrir diferents fils de discussió i que cadascun es diferencies de la resta. També m'agrada posar diferents categories a les contribucions, encara que crec que de vegades no sabem exactament a quina categoria incloure la nostra contribució. Potser son massa rígides.

2. Com creus que la nova eina ha incidit en el grau quantitatiu de la teva participació. (valora 0 -5 i fes els comentaris que creguis convenients)

   DOS - El problema de que hagi tants fils diferents i amb tantes contribucions fa que de vegades es repeteixen opinions i línees de debat i fa una mica més difícil obrir un debat nou o contribuir als ja existents. A més no he trobat com veure només les meves aportacions o saber a quin fils s'han fet.

3. Com creus que la nova eina ha incidit en la qualitat de les teves aportacions. (valora 0 -5 i fes els comentaris que creguis convenients)

   TRES - Es pot llegir en una mateixa pantalla totes les contribucions i això fa que puguis seguir la línea de pensament dels companys i fer l'aportació pròpia. Per altra part el fet de no tenir cap classificació prèvia no t'aporta ninguna informació que t'ajudi a triar en quin fil vols contribuir. Sé que el més adequat seria llegir el inici de tots el fils, però quan hi ha tants fils si, a més, les contribucions posteriors son extenses es difícil seguir tots els fils.

4. En comparació als espais de fòrum i debat de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   El principal avantatge per mi està a la presentació dels fils. Com ja he dit abans tens una visió molt més general que al espai de l'aula. Si intentéssim posar tots els missatges de l'espai a l'aula seria molt poc manejable. A més cadascun de nosaltres pot obrir el seu propi fil i tancar-lo quan ho consideri oportú. Al fòrum no es pot fer. El principal inconvenient, per mi es que tens poca informació anterior a l'obertura de les diferents contribucions. Tens la categorització, però no hi ha un encapçalament que t'avanci el contingut i t'ajudi a triar d'una forma més ràpida. El tema i les carpetes del fòrum faria aquesta feina.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Més que a l'eina potser jo modificaria la manera de fer el debat. Personalment crec que seria interessant que hagués una mena de moderador que triés unes quantes línees de debat i partir d'aquestes per fer les nostres aportacions. Fins i tot es podria considerar fer una aportació a cada tema per garantir una millor participació o permetre als estudiants obrir els seus propis temes. Això requeriria que l'eina permeti la figura del moderador. Per altra banda seria interessant poder veure on has fet les teves aportacions d'un cop d'ull per poder continuar participant en una determinada línea.

Questionnaire #4

Preguntes sobre la nova eina Fòrum de Discussió

1. Com valores en general la nova eina Fòrum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   Valoració 3. Crec que encara li falta uns quants jocs de proves més.

2. Com creus que la nova eina ha incidit en el grau quantitatiu de la teva participació. (valora 0 -5 i fes els comentaris que creguis convenients)

   Valoració 4. La veritat que és més fàcil el seguiment amb aquesta eina abans que en el fòrum en l'espai de l'estudiant.

3. Com creus que la nova eina ha incidit en la qualitat de les teves aportacions. (valora 0 -5 i fes els comentaris que creguis convenients)

   Valoració 2. Crec que no influeix en la qualitat de les aportacions.

4. En comparació als espais de fòrum i debat de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Comparant amb l'espai del fòrum de l'aula, aquesta eina és molt millor. El debat en el fòrum de l'aula és molt difícil de fer un bon seguiment. Per mi, crec que s'hauria d'eliminar el fòrum de l'aula i fer servir una eina com aquesta, un cop sigui més robusta.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Aquestes serien les millores o Poder triar el nombre de fils per pàgina; o Donar per llegit un fil un cop entris (com a la UOC); o Filtrat de fils: ordenar per data, llegits, no llegits...; o Poder tornar enrere un cop hagis llegit un fil; o Menys opcions a l'hora d'introduir una resposta; o Ordre invers d'aportacions. Començar per la primera i tenir les respostes cap avall, com un fòrum qualsevol;

   En resum, crec que s'hauria de triar aquesta eina per fer els debats.

Questionnaire #5

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   Como nota le pondré un 4. Al principio ha costado un poco el saber como realizar las aportaciones y como catalogarlas, al igual que acabar de entender lo de dejar debate abierto o cerrado. En general me ha parecido una herramienta ordenada y clara para el seguimiento del debate. Tal vez, si no hubiera empezado el debate aportando cada usuario su solución inicial hubiera sido todavía un poco más cómodo de seguir.

2. Com creus que la nova eina ha incidit en el grau quantitatiu de la teva participació.

   (valora 0 -5 i fes els comentaris que creguis convenients)

   La nueva herramienta facilita el seguimiento del debate mostrando de forma más ordenada las aportaciones pendientes del resto de los compañeros. Le pongo un 4.

3. Com creus que la nova eina ha incidit en la qualitat de les teves aportacions. (valora 0 -5 i fes els comentaris que creguis convenients)

   Desde mi punto de vista, la herramienta que se utiliza no me ha influido en la calidad de las aportaciones ya que éstas dependen del material que se esté debatiendo y de la calidad de las aportaciones del resto de los compañeros. Eso sí, facilita en gran medida el seguimiento del debate y la localización de las aportaciones pendientes de leer.

4. En comparació als espais de fòrum i debat de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Como he contestado anteriormente, la nueva herramienta ayuda al seguimiento del debate, tanto en la localización de las distintas líneas de debate como en la localización de aportaciones pendientes de leer.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Tal vez una de las aportaciones que haría a la herramienta sería la de incorporar de alguna manera algún tipo de aviso en el caso de que alguien conteste alguna de las aportaciones que haya realizado uno mismo, ya sea mediante avisos nada más entrar en la aplicación como la de mandar algún mail avisando de dichas repuestas.

Questionnaire #6

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   5. Una gran idea.

2. Com creus que la nova eina ha incidit en el grau quantitatiu de la teva participació. (valora 0 -5 i fes els comentaris que creguis convenients)

   3. M'he sentit més motivat a participar.

3. Com creus que la nova eina ha incidit en la qualitat de les teves aportacions. (valora 0 -5 i fes els comentaris que creguis convenients)

   2. Un pèl massa de dispersió. Massa fils.

4. En comparació als espais de fòrum i debat de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   És més fàcil seguir en quins fils has participat i en quins no. Això permet eliminar fils en els quals no has cregut que puguis aportar res de nou.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Una búsqueda per aportació (llargada, títol, tipus d'aportació,etc.). En certs moments era difícil saber quin era aquell fil en el que havies introduit un comentari al qual volies veure les reaccions, sense saber el número.

Questionnaire #7

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   4- Tenint en compte que està en fase de proves, crec que és una molt bona eina per a gestionar debats, tot i que hi ha hagut problemes de connexió i falta d'experiència per part dels participants.

2. Com creus que la nova eina ha incidit en el grau quantitatiu de la teva participació. (valora 0 -5 i fes els comentaris que creguis convenients)

   2- No crec que la nova eina hagi influït en el meu grau de participació.

3. Com creus que la nova eina ha incidit en la qualitat de les teves aportacions. (valora 0 -5 i fes els comentaris que creguis convenients)

   3 - Al contrari de la resposta anterior si que crec que la qualitat de les respostes pot ser millor en aquest entorn degut a la major facilitat de seguir els diàlegs i de poder diferenciar els temes els quals es tracta.

4. En comparació als espais de fòrum i debat de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   La nova eina fa que es pugi seguir millor el debat i poder tenir diferenciats els diferents diàlegs que es poden portar a terme. Però això també provoca que un mateix tema es pugi repetir en molts debats diferents, i per tant les aportacions siguin repetitives. En el nou espai hi ha hagut els problemes típics d'una eina en fase de proves, exceptuant això crec que és molt correcte.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Crec que l'eina és molt adequada per a la realització de debats.

Questionnaire #8

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   Valoració: 4 Comentaris: fàcil, intuïtiva, problemes de servei comprensibles. Vaig enviar un missatge massa llarg i va tirar endavant en comptes d'anular la publicació!

2. Com creus que la nova eina ha incidit en el grau quantitatiu de la teva participació. (valora 0 -5 i fes els comentaris que creguis convenients)

   Valoració: 4 Comentaris: és fàcil navegar per tots els fils i participar-hi. Bona informació sobre les participacions individuals en cada fil.

3. Com creus que la nova eina ha incidit en la qualitat de les teves aportacions. (valora 0 -5 i fes els comentaris que creguis convenients)

   Valoració: 4 Comentaris: la creació de fils permet acotar i agilitar un debat que d'altra manera seria massa dispers.

4. En comparació als espais de fòrum i debat de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Millores i avantatges: permet crear fàcilment un nou fil (amb permisos per tancar-lo). Inconvenients: l'arbre de navegació és més complet als espais de l'aula (permet una visió simultània del llistat de fils i el contingut d'un missatge).

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Opció de previsualitzar els missatges abans de publicar-los. Permalinks als missatges per poder citar-los entre fils diferents. Clàssica col·lecció de botonets d'edició (copiar, pegar, cursiva, link, etc) com al correu de la UOC.

Questionnaire #9

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   [3] Crec que encara ha de millorar una mica, no parlem dels petits defectes dels signes de puntuació o de les fallades de connexions que de ben segur es solucionaran, sinó que el que trobo una mica lleig es el sistema de navegació entre les carpetes, els fils, i els comentaris. Trobo que s'hauria d'implementar un sistema de navegació més jeràrquic i eficient que permeti moure's al detall o l'engròs dins els diferents nivells d'informació. Cal tenir en compte un debat aparentment senzill com el que hem realitzat ha generat centenars de missatges en diferents fils amb diferents tipificacions, etc.

2. Com creus que la nova eina ha incidit en el grau quantitatiu de la teva participació. (valora 0 -5 i fes els comentaris que creguis convenients)

[3] No crec que la eina tingui influencia respecte a la quantitat d'aportacions, encara que ara per ara trobo més proper el fòrum tradicional, doncs permet una visió general dels missatges i una lectura de 'tot' que a la nova eina no hi es.

3. Com creus que la nova eina ha incidit en la qualitat de les teves aportacions. (valora 0 -5 i fes els comentaris que creguis convenients)

   [4] Respecte a la qualitat si que crec que ha millorat senzillament per la tipificació que es pot aplicar a les aportacions, així queda clar si estem obrint o tancant un debat o fent una pregunta etc.

4. En comparació als espais de fòrum i debat de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Avantatges:
   la separació de les carpetes
   la separació dels fils de discussió
   la tipificació que es pot realitzar als missatges
   la informació de si un mateix ha aportat una entrada o no dins un fil de discussió

   Desavantatges:
   La pèrdua de la visió de conjunt
   La complicació de capçalera, abans hi havia nomes el emissor i el títol
   La complicació de la lectura dels missatges, s'ha d'entrar d'un en un, no es pot llegir 'tot el que encara no he llegit de tots els fils
   El que no estigui integrat al campus de la UOC i tenir que introduir una contrasenya aliena

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Trobo a faltar un sistema de notificació per subscripció dels fils interessants, sobre tot els creats per un mateix, per tal d'atendre amb rapidesa les qüestions que es puguin plantejar. Addicionalment es podria mantenir una traça de les aportacions a cadascun dels fils. També deixaria els missatges en el seu ordre de lectura cronològic, i no com ara que es ordre invers.

Questionnaire #10

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   1. Independentment dels problemes tècnics que han sorgit al utilitzar-la, es una eina bastant caòtica. Es complicat seguir els fils.

2. Com creus que la nova eina ha incidit en el grau quantitatiu de la teva participació. (valora 0 -5 i fes els comentaris que creguis convenients) 4. Els problemes tècnics i el fet de ser una eina poc còmoda per seguir els debats, ha incidit de forma rellevant en les meves aportacions.

3. Com creus que la nova eina ha incidit en la qualitat de les teves aportacions. (valora 0 -5 i fes els comentaris que creguis convenients) 3. No hi ha hagut gaire temps per estudiar totes les aportacions degut als continus problemes de l'eina

4. En comparació als espais de fòrum i debat de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Avantatges. Poder categoritzar les aportacions i saber el grau de participació de l'alumne Problemes. Incovenients. Difícil seguir els fils de les converses, encara que el forum i debat de l'aula no presenta gaires avantatges respecte aquesta eina.

Questionnaire #11

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   És una bona eina, millor que l'actual al Campus Virtual. Però cal depurar el seu plantejament. Ha esta molt bona la idea de crear una carpeta per fer proves i un fil per a dubtes de debat. Ha estat, per la meva opinió, poc encertat encetar el debat amb més de 40 fils.

   Jo hagués creat els fils corresponents a punts concrets del debat. Si es planteja un tema es poden fer diferents preguntes que creïn posicions enfrontades. I per cada pregunta un fil de discussió. És a la vegada més fàcil de moderar.

2. Com creus que la nova eina ha incidit en el grau quantitatiu de la teva participació. (valora 0 -5 i fes els comentaris que creguis convenients)

   No valoraria aquesta resposta. La nova eina no ha fet variar el meu grau de participació, no ha fet que tingués més ganes de participar. He participat el que el temps lliure que tinc em permet.

3. Com creus que la nova eina ha incidit en la qualitat de les teves aportacions. (valora 0 -5 i fes els comentaris que creguis convenients)

   La eina no fa que es participi amb més qualitat, ha de facilitar l'accés al debat. Tampoc tinc valoració per aquesta pregunta.

4. En comparació als espais de fòrum i debat de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Hi ha molts avantatges, es creen diferents fils de debat (que ajuda a estructurar el debat). Però un altra vegada penso que han estat excessius fils oberts a l'hora. Menys fils de debat ajuda a tenir una visió més compacte del debat i ajuda a moderar-lo millor.

   Si es creen fils per a cadascun dels participants es pot caure molt en la repetició d'arguments. És millor posar una quota de participació (tipus un mínim d'aportacions al debat) i espaiades en el temps, no totes al principi o al final per cobrir la quota.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Si vols aportar alguna cosa a un fil (diàleg) voldria tenir a la vista totes les contribucions realitzades fins el moment, no només la darrera aportació. No sempre vols respondre la darrera contribució, fins i tot vols revisar què havies dit en una contribució anterior.

   I sento ser una mica repetitiu però he trobat massa fils de debat oberts a l'hora.

   L'enquesta final es pot fer en un formulari tipus plana web en comptes d'un document escrit. Dóna més agilitat a l'hora de fer les valoracions i crec que hauria de ser anònima.

6. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Seria convenient:

   - Rebre un e-mail quan algú contesta a una aportació teva

   - Integrar l'eina dins de l'aula

   - Veure les últimes respostes a fils on has participat

   - Canviar el sistema de paginació

   - Cercador

   Crec que tant aquesta eina com les de l'aula no faciliten la comunicació entre els alumnes. Caldria treballar més en aquest aspecte donat que son les eines més importants per comunicar-nos.

Questionnaire #12

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   Valoració: 3 Comentaris: M'he fixat que al principi té masses opcions en el moment d'escollir el tipus de contribució al fil (INFORMACIÓ-Petició, INFORMACIÓ-Ampliació, ...) que m'ha donat confussió per escollir-lo ja que en una mateixa resposta pot haver-hi una petició, una ampliació o salutació i no crec pas que per cadascuna hagi de crear un missatge independent per cadascuna de la resta. També he vist que poden existir dintre d'un mateix fil diversos fronts oberts i cadascun de temàtica diferent amb el que en el moment de tancar cadascún tenia que resumir-ho de forma mesclada ja que no puc "moderar" el meu fil.

2. Com creus que la nova eina ha incidit en el grau quantitatiu de la teva participació. (valora 0 -5 i fes els comentaris que creguis convenients)

   Valoració: 3 Comentaris: En veure que la gent ha contestat en el meu fil, m'ha donat peu a poder replicar, ampliar o discutir les seves aportacions i també a donar les gràcies a les seves crítiques. Al saber també quantitativament les aportacions de la resta de companys al debat, em fa veure quin és el grau mig de les aportacions dels alumnes de manera que jo em vulgui superar aportant més (tot i que com se sol dir, més val la qualitat que la quantitat). El problema ha estat potser les dates en què s'ha posat en marxa el debat, és a dir, les vacances de Setmana Santa i els possibles errors generats per l'eina, hauràn fet que el número d'intervencions no hagi estat el millor.

3. Com creus que la nova eina ha incidit en la qualitat de les teves aportacions. (valora 0 -5 i fes els comentaris que creguis convenients)

   Valoració: 2 Comentaris: L'eina per si mateixa, no fa que les respostes siguin millors o pitjors, això depèn de la persona qui escriu. Tot i així he vist que la caixa de text on s'escriu és molt petita donant peu a voler sintetitzar massa les possibles respostes (no s'escriu tot el que es vol) de manera que la qualitat baixi. No obstant, quan encara estàs descobrint com funciona l'eina les respostes són encara "de proves" però un cop saps com funciona et concentres en l'aportació i per tant creix la qualitat.

4. En comparació als espais de fòrum i debat de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

Aquesta nova eina aporta un valor afegit al servei de que disposa la UOC enfront a donar una qualitat de servei. Només seria dessitjable que s'estandaritzés el seu ús per a moltes més assignatures que tinguéssin debats. Si després de tot no s'utilitza més, crec que haurà estat una pèrdua de temps per a la meva persona (pel cost de l'aprenentatge) tot i que per la Universitat li haurà servit per saber si ha estat bona o dolenta aquesta prova pilot. Les avantatges sobre el sistema tradicional de l'espai de l'aula és la seva dedicació, és a dir, que és un programari a mida per fer debats. Per tant al ser-hi dedicat les seves opcions, versatilitat, ús són millors que l'altre. Com a tema important ha estat la categorització dels missatges així com crear o tancar fils. Els inconvenients són l'aprenentatge d'un nou sistema i els errors que el programari nou dóna (tot i que s'aanirà depurant).

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Ara per ara, no he vist que pogués configurar al meu gust l'eina, ja sigui per canviar colors o canviar la contrasenya. S'hauria d'integrar amb el campus virtual de manera que no tornés a demanar un altre cop l'usuari/contrasenya. Els accessos de la part de l'esquerra de "tens novetats" podrien ser linkables per anar directes al fil... He vist que s'ha posat paginació en els fils de discussió (cosa que s'agraeix per no tenir que anar veient cada cop els 5 següents per anar a l'últim fil). La paginació s'hauria de pulir, ja que si existeixen moltíssims més fils no hi hauria lloc per posar tots els rangs de 5 en 5. Es podria posar el típic anar al primer o últim, següent pàgina, última pàgina, etc. i/o que nosaltres escollim el número de fils per pàgina. Com tema més important caldria que tingués un cercador per veure les opinions que hi ha en un debat segons els criteris de cerca.

Questionnaire #13

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   Valor: 4.

2. Com creus que la nova eina ha incidit en el grau quantitatiu de la teva participació. (valora 0 -5 i fes els comentaris que creguis convenients)

   Valor: 4. Es motivador el dato estadístico de participación.

3. Com creus que la nova eina ha incidit en la qualitat de les teves aportacions. (valora 0 -5 i fes els comentaris que creguis convenients)

   Valor: 3. El disponer de hilos de discusión independientes, focaliza la calidad de las aportaciones por una mayor especialización.

4. En comparació als espais de fòrum i debat de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Mejoras: Entorno especializado en los debates y facilidad de seguimiento de los hilos de discusión. Muy positivo el seguimiento estadístico de la participación y la valoración de la calidad de las intervenciones.

   Inconvenientes: Necesitar una acceso independiente al aula. Se podría poner un enlace directo desde el aula, sin necesidad de tener que autenticarse de nuevo. Al salir del hilo de discusión se vuelve a la página inicial, siendo positivo volver a la página que se está trabajando.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Camentadas en el anterior punto.

Questionnaire #14

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Fòrum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   Tal com he comentat en l'apartat anterior, a part d'alguns moments inicials en que l'eina es penjava i que la navegació era de fil en fil, que s'ha millorat en els dies posteriors, jo he pogut treballar amb normalitat. Es cert que els primers dies m'ha costat habituar-me a l'eina però un cop llegit el correu enviat al fòrum de l'assignatura ja he entès el seu funcionament.

2. Com creus que la nova eina ha incidit en el grau quantitatiu de la teva participació. (valora 0 -5 i fes els comentaris que creguis convenients)

   El fet de tenir un debat a part de la PAC fa que es creï una dependència, quasi diària, amb l'assignatura, ja que en qüestió d'un o dos dies les aportacions fetes pels estudiants han vairat moltíssim. La PAC es un exercici autoformatiu, on saps el termini i et pots organitzar el treball i la feina per acomplir amb el dia d'entrega. Això no es pot dir amb el debat ja que depens del moment lliure de l'altra gent i per això es crea la dependència que he esmentat anteriorment.

3. Com creus que la nova eina ha incidit en la qualitat de les teves aportacions. (valora 0 -5 i fes els comentaris que creguis convenients)

   No crec que m'hagi influït l'eina amb la qualitat de les meves aportacions.

4. En comparació als espais de fòrum i debat de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   No tinc aportacions per aquesta pregunta

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   No tinc aportacions per aquesta pregunta

Questionnaire #15

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   Valoració 4. Al principi una mica feixuga, fins que vam tenir la possibilitat de moure'ns per pàgines. L'ús dels colors per determinar a quins fils hem intervingut esta molt be

2. Com creus que la nova eina ha incidit en el grau quantitatiu de la teva participació. (valora 0 -5 i fes els comentaris que creguis convenients)

   Valoració 4. És molt senzill poder contestar a la resta de companys. Realment no m'ha suposat cap impediment l'eina alhora de participar.

3. Com creus que la nova eina ha incidit en la qualitat de les teves aportacions. (valora 0 -5 i fes els comentaris que creguis convenients)

   No veig la relació entre la qualitat de les meves aportacions, i l'eina. Valoració 0

Table 6.2: Main statistics extracted from the second experience using both discussion tools.

Statistics	**ST**	**DF**
Number of students	66	66
Number of threads	52	31
Total of posts	229	417
Mean number (posts/thread)	M=2.2 SD=4.4	M=13.4 SD=6.1
Mean number (posts/student)	M=3.4 SD=2.1	M=6.3 SD=5.2

4. En comparació als espais de fòrum i debat de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Es nota que l'eina esta pensada per aquest tipus de forum de dicusions. És més àgil, i senzilla de participar.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Que fós possible de rebre per e-mail quan algú ha contestat la meva aportació. Només que digués que hi ha hagut canvis, ja n'hi hauria prou

Experience #2 - Fall term 2007

- Course: Management of Organizations and Computer Science Projects (GOPI)
- Assignment type and duration: Class discussion activity for 20 days
- Participation type: Required
- Discussion tools: UOC's *standard tool* (ST) and *Discussion Forum* (DF) version 0.8 (prototype)
- Potential participants: 66 (ST) + 66 (DF) undergraduate students

Statistic results

See Table 6.2.

Questionnaire results

Participation (optional): 7/66 (10%). Most relevant questionnaires' responses (Original language):

Questionnaire #1

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Fòrum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   3, a banda del temps de resposta, que un pot assimilar, no es gaire intuïtiva. Potser hauria de millorar la interfície del usuari

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

Potser positivament (4) , donat que al no ser fàcil de entendre, en un principi entraves més del compte al fòrum, i sense voler t'anaves embolicant en el debat

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Ídem anterior (4)

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Avantatges: la possibilitat de donar un enfocament a la teva intervenció es molt important: ja d'entrada dius si estàs d'acord(AGREE) o no (DISAGREE) amb el company. Això facilita la lectura donat que pots fer una selecció: els que no estan d'acord amb tu son els primers que llegeixes, els altres els deixes per el final

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Ajuda on-line Entorn més amigable en general

Questionnaire #2

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   4 La eina és bona, àgil, agradable i molt navegable. Llàstima per la velocitat.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   2 en quantitat i 3 en qualitat És difícil de explicar però, en algun moment entrar en l'aplicació era com anar al dentista. Només de pensar ja et fa mal tot. Crec recordar que alguna vegada he estat fins a tres hores per poder mal llegir els comentaris. Si hagués tingut més velocitat hagués participat més vegades.

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   2 en quantitat i 3 en qualitat Crec que a tothom li ha passat una mica el mateix

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Poder veure les notes que posava el consultor ha estat tota una experiència. Durant el debat entre 5 i 6 persones han estat valorades amb una B. Això fa que ràpidament vagis a veure que han escrit els teus companys per veure com poder millorar. En el meu cas no ha estat possible pujar el nivell per falta de temps però ha anat bé per veure que valora el consultor

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Indudablement velocitat. Desprès he trobat a faltar ordre amb les participacions i la impossibilitat de poder imprimir. També ha estat una experiència el valorar les aportacions dels companys. De fet crec que no estic/estem preparat per fer-ho

Questionnaire #3

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   L'estructura és molt bona i ajuda a la participació, però el seu funcionament és exasperant degut a la seva extrema lentitud que pot arribar al col·lapse en determinats moments on has de deixar la participació degut a la impossibilitat d'accedir al debats. Ahir tarda/vespre (dia 30 de Octubre, entre les 16:00 i les 24:00 hores) no vaig poder avaluar, ni donar per llegit ni respondre a cap company. D'acord que era el darrer dia, però la meva participació ha estat molt regular al llarg del període de debat i m'he trobat en la mateixa situació moltes vegades. Valoració 1.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Com que per mi ha estat una eina atractiva crec que ha afavorit de forma molt positiva. Valoració 5

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   El format és atractiu i afavoreix la competitivitat entre els alumnes per millorar la quantitat d'intervencions (Valoració 5), altra cosa és la qualitat donat que no és possible comparar ni valorar adequadament intervencions molt ben estructurades amb altres d'assentiment o molt curtes (Valoració 3)

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Millor estructura, millor separació per temes, avaluació de la teva participació en comparació a la resta d'alumnes, possibilitat d'obrir nous debats de forma fàcil

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Imprescindible: Ha de ser més ràpida.

Questionnaire #4

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   5 . Valoro molt positivament aquesta eina i quan s'hagin realitzat les modificacions pertinents, crec que pot ser molt útil per a la resta d'assignatures de la UOC.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   5

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

4. L'eina facilitava entrar en qualsevol moment i si disposaves de poc temps, facilitava la tasca de centrar-te en un tema de discussió sense haver de valorar per el camp ”Asunto” si t'interessava obrir el missatge i llegir-lo o no (com passa als debats tradicionals).

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Crec que la nova eina té molts avantatges respecte al debat tradicional perquè et permet organitzar més bé la informació en distribuir-se els temes en fils. Facilita la participació perquè només cal que triïs els fils que més t'interessen i anar-los seguint. Amb el debat tradicional hi havia una feina ingent per seguir els temes de debat i acabaves imprimint totes les col·laboracions dels companys (que fàcilment eren un centenar de missatges) per poder seguir bé els fils (i imprimies missatges de temes que realment no t'interessava seguir), escrivint les teves aportacions a mà i després traspassant-les al debat.

   Val a dir que també hi ha els inconvenients que no s'han cansat de comentar els companys: lentitud, el fet de que cada vegada que avalues la utilitat, assenteixes a una contribució o dónes per llegit, et torni a la capçalera de la pàgina. Com vaig comentar en un missatge al fòrum, penso que el ”dóna per llegit” s'acceleraria si es fes amb un quadre de validació (com es fa al fòrum de l'aula) .

   En el meu cas particular també se m'ha fet complex saber quina etiqueta havia de posar al missatge i no he acabat d'entendre la utilitat dels missatges d'ànim en el debat.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Velocitat. Modificar les etiquetes dels missatges per fer-ho més intuïtiu. Fer un manual una mica més entenedor (es fa una mica feixuc entendre el manual que vareu facilitar abans de començar el debat).

Questionnaire #5

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   1 Veure tants missatges pendents t'estressa nomes obrir. Es impossible donar tot per llegit. Sempre que no estàs d'acord amb alguna afirmació has de respondre i possiblement el fil ja es massa ple de respostes com la que tu vols expressar i al final passes per no repetir un altre vegada lo mateix

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   0. Quantitat. Era realment dificultós estar al dia. Hi havia fils que donaven error al entrar. Al donar per llegit un missatge podia tardar un minut a tornar a veure el fil. Fa que no pugis avaluar els missatges dels companys com mereixen. 2. Qualitat. No saps quin tipus de missatge enviar, a missatges que voldries respondre no pots perquè son d'un altre tipus. Això si, permet comparar les teves aportacions amb la dels altres companys, el que et fa espavilar una mica

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

3. El forúm permet que la gent participi mes que en altres assignatures. Però la seva lentitud desespera i fa que desisteixis mes d'una vegada. Vas donant per llegits missatges i fas un altre cosa mentre esperes i fins i tot arriba un moment que t'en oblides del que estaves llegint.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Valorar les opinions del companys. Comparar les aportacions. Llegir tots els missatges d'una tacada afavoreix llegir-ho tot. El temps per donar per llegit un missatge o el que t'obligui a respondre si no estàs d'acord fa que passis de flags i de pendents i llegeixis però no marquis

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Que no peti. Rapides al donar per llegir : marcar els missatges i donar al botó corresponent. Tipus de missatges mes entenedors i apropiats i indicar un desacord encara que no vulguem donar una resposta(de vegades el que vols dir ja esta expressat i es repetitiu)

Questionnaire #6

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Fòrum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   En general, es una eina molt intel·ligent que intenta una major interacció entre els estudiants i l'assignatura i de fet , ho aconsegueix positivament, però com tota nova eina hi ha mancances tècniques que puntualment , influeixen molt negativament. Per aquet motiu, la meva valoració es un 3.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Des de el meu punt de vista, ha de incidir en el mateix grau però sempre sota el control d'un moderador. Una valoració general de 4 en tots dos sentits

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Fent una revisió de les estadístiques, hi ha una gran participació de nivell, en general de molta interacció i per això posaria un 5 a tots dos punts.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Una gran avantatge del producte esta en la raó que es pot extrapolar a qualssevol de les assignatures de la Carrera Universitària i aquesta reutilitazacio es molt positiva al debat tradicional. Conceptualment es una bona eina que tindria que evolucionar tècnicament.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

Table 6.3: Main statistics results from the third experience using both discussion tools.

Statistics	**ST**	**CoLPE**
Number of students	43	43
Number of threads	29	17
Total of posts	174	93
Mean number (posts/thread)	M=6.0 SD=2.7	M=5.5 SD=4.5
Mean number (posts/student)	M=4.0 SD=1.6	M=2.2 SD=3.8

Una millora significativa seria afegir rapidesa i dinamisme gràfic similar a un sistema de correu convencional , alhora d'incorporant algun petit apartat favorits per "deixar" marcats o emmagatzemar aquelles aportacions que , sempre des d'un punt de vista personal, siguin mes adients (en 400 o 500 aportacions es fàcil perdre el fil entre tot un bosc de missatges).

Experience #3 - Fall term 2007

- Course: Course: Methodology and Management of Computer Science Projects (MGPI)
- Assignment type and duration: Class discussion activity for 14 days
- Participation type: Required
- Discussion tools: UOC's *standard tool* (ST) and *Communities of Learning Practice Environment* (CoLPE) version 0.6 (prototype)
- Potential participants: 43 (ST) + 43 (CoLPE) graduate students

Statistic results

See Table 6.3. See also subsection 5.4.2 in Chapter 5 for a detailed description and analysis of the technical issues observed while running this experience, which influenced it negatively.

Questionnaire results

Participation (optional): 7/43 (16%). Most relevant questionnaires' responses (Original language):

Questionnaire #1

Preguntes sobre la nova eina CoLPE

1. Com valores en general la nova eina CoLPE per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   5, ET PERMET FER UN SEGUIMENT PER TEMES, TROBO MOLT UTIL VEURE ELS MISSATGES QUE NO HAS LLEGIT, TAMBÉ FACILITA MOLT L'IMPRIMIR ELS TEMES DELS DEBATS, AIXO HO CONSIDERO UN GRAN AVENÇ RESPECTE EL METODE ANTERIOR, JA QUE POTS IMPRIMIR PER TEMES

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

3

3. Quin paper creus que ha jugat la nova eina en la participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   4

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula? Si us plau, proporciona crítiques constructives que ajudin a millorar l'entorn.

   IMPRESSIO MISSATGES, POTSER LLETRA UNA MICA GRAN, PERÒ ET PERMET FER PARRAF SABENT EL SEU TAMANY

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants en general. Quines funcionalitats creus que no són necessàries per portar a terme un debat com aquest.

   ELS TIPUS INFORM NO VEIG MOLT LA SEVA UTILITAT, CALDRIA UN BON MANUAL D'AJUDA DEL ENTORN ...

Questionnaire #2

Preguntes sobre la nova eina CoLPE

1. Com valores en general la nova eina CoLPE per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   3. L'eina proposava per a fer el debat ha estat fàcil d'entendre i fer-la anar. A més a més és més fàcil seguir els fils de cada aportació.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Grau quantitatiu. 2. atès que al ser una nova eina, ha estat al principi una mica poc entenedor fer les aportacions adequades. Grau qualitatiu. 4. penso que a nivell qualitatiu s'ha guanyat més que fent-los de la manera antiga el debat.

3. Quin paper creus que ha jugat la nova eina en la participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   3. Penso que aquesta eina es pot fer servir per fer debats sense cap inconvenient atès que aporta totes les eines necessàries per fer-lo correctament.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula? Si us plau, proporciona crítiques constructives que ajudin a millorar l'entorn.

   L'home per molt que ens pesi, és animal de costums i atès que a nivell personal porto ja sis anys fent debats a la uoc amb l'eina tradicional, estic més acostumat a fer-los com abans. No obstant es qüestió de temps i no costaria gaire utilitzar aquesta nova eina.

   Una millora important que té és que utilitzes un espai fora del campus i evites que altres tasques influeixin a l'hora de fer el debat, és a dir, qu7an accedeixes al debat estàs pel debat, en el campus pots fer vàries coses que no tenen a veure amb el debat pròpiament dit.

Un inconvenient important que m'he trobat és que amb l'anterior eina de debat, podries tenir amb una sola pàgina tot el contingut del debat per llegir-lo off-line i guardar-te'l en disc. Amb la nova eina aquesta funció no està disponible.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants en general. Quines funcionalitats creus que no són necessàries per portar a terme un debat com aquest.

   Per mi l'eina ha estat suficient per portar el debat. No he tingut gaires problemes per participar-hi (traient dels primers dies que crec que va caure un parell de vegades el servidor i no deixava accedir-hi).

Questionnaire #3

Preguntes sobre la nova eina CoLPE

1. Com valores en general la nova eina CoLPE per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   La valoro amb un 3. Crec que li queda molt per millorar. Per exemple seria molt útil que les aportacions d'un fil es poguessin llegir sense haver de clickar cada vegada a l'aportació que vols llegir. Com si fos un fòrum.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   No considero que hagi incidit l'eina en el meu grau de participació. La informació era accessible i la eina és fàcil d'utilitzar. Simplement falta millorar la usabilitat donat que actualment considero que calen massa clicks per realitzar les tasques. És a dir, passes més temps clickant aquí i allà que llegint/escribint pròpiament la informació, quan és aquesta última el que realment interessa.

3. Quin paper creus que ha jugat la nova eina en la participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Al igual que abans, no considero que hagi incidit l'eina en el grau de participació.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula? Si us plau, proporciona crítiques constructives que ajudin a millorar l'entorn.

   Crec que és molt positiu el fet de poder llegir la informació de manera molt més ordenada i relacionada. La resta ja ho he comentat a les preguntes anteriors.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants en general. Quines funcionalitats creus que no són necessàries per portar a terme un debat com aquest.

   No response.

Questionnaire #4

Preguntes sobre la nova eina CoLPE

1. Com valores en general la nova eina CoLPE per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

Puntuació: 4 Comentari: M'hi he sentit còmode. Al principi funcionava lentament. Cal que l'eina sigui ràpida.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Puntuació: 0 Comentari: No he trobat cap diferència quantitativa ni qualitativa respecte a l'ús del fòrum tradicional.

3. Quin paper creus que ha jugat la nova eina en la participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Puntuació: 0 Comentari: No he trobat cap diferència quantitativa ni qualitativa respecte a l'ús del fòrum tradicional.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula? Si us plau, proporciona crítiques constructives que ajudin a millorar l'entorn.

   Crec que l'eina aporta més informació al consultor que no pas a l'alumne. No veig massa canvis de cara a l'alumne respecte als fòrums tradicionals.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants en general. Quines funcionalitats creus que no són necessàries per portar a terme un debat com aquest.

   Trobo a faltar una opció per accedir als missatges no llegits de forma ràpida. Actualment feia clic sobre el fil i després sobre "tots els missatges" i finalment feia scroll per veure els nous. Crec que són massa passes. Seria bo fer-ho des de la pantalla on es veuen tots els fils amb un enllaç que digues "Veure comentaris nous".

   Pregunto: Per què no s'ha fet l'autoavaluació amb Colpe en comptes d'utilitzar aquest document?

Questionnaire #5

Preguntes sobre la nova eina CoLPE

1. Com valores en general la nova eina CoLPE per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   (3) De fet crec que particularment no li he tret tot el profit ja que m'he limitat a utilitzar el mínim necessari per poder participar i llegir la informació.

   Crec que és una eina adequada (ja que esta pensada amb aquest propòsit) però estaria bé conèixer la part de 'backend' on és deuen de veure les estadístiques de participació etc..

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   (1) La meva participació hauria estat la mateixa amb o sense aquesta eina. No ha aportat cap efecte especialment motivador .

3. Quin paper creus que ha jugat la nova eina en la participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   (1) Mateix comentari que en el punt anterior.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula? Si us plau, proporciona crítiques constructives que ajudin a millorar l'entorn.

Table 6.4: Main statistics results from the fourth experience using both discussion tools.

Statistics	**ST**	**DF**
Number of students	35	34
Number of threads	54	12
Total of posts	109	214
Mean number (posts/thread)	M=2.0 SD=1.8	M=17.8 SD=28.6
Mean number (posts/student)	M=3.1 SD=1.1	M=6.2 SD=2.7

El que més m'ha agradat és que permeti la categorització del tipus d'intervenció, la notificació de noves aportacions sobre cada fil. El que menys, la seva interfície d'usuari, poc còmoda i intuïtiva.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants en general. Quines funcionalitats creus que no són necessàries per portar a terme un debat com aquest.

   Millorar la cerca, saber qui ha llegit el teus comentaris, presentar la informació de formes diferents, simplificar la interfície (o disposar de diferents nivells o graus d'utilització)

Experience #4 - Spring term 2008

- Course: Management of Organizations and Computer Science Projects (GOPI)
- Assignment type and duration: Class discussion activity for 20 days
- Participation type: Required
- Discussion tools: UOC's *standard tool* (ST) and *Discussion Forum* (DF) version 0.8 (prototype)
- Potential participants: 35 (ST) + 34 (DF) undergraduate students

Statistic results

See Table 6.4[1].

Questionnaire results

Participation (optional): 8/34 (23%). Most relevant questionnaires' responses (Original language):

Questionnaire #1

Preguntes sobre la nova eina Discussion Forum

1. . Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   Un 3. Està molt bé com a iniciativa, però de vegades és un pèl confús el sistema de rèpliques, contrarèpliques, obertura de fils nous, establir si estàs d'acord o no amb alguna de les intervencions, etc.

[1]Note this very high SD statistic in the posts/students mean. This was chiefly caused by two empty discussion threads, which were open by students just before closing the discussion. Without this outlier, SD=5.4.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Quantitatiu i qualitatiu: 3. No hi ha influït decisivament en el meu cas, però sí que ha estat engrescador participar en el debat. Quantitativament és una qüestió de temps, en el meu cas, i qualitativament no hi puc fer més.

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Alta participació: 4. Com ja he comentat crec que és atractiu participar en el debat, ja que els propis comentaris dels alumnes comporten les ganes de dir-hi la teva. Qualitat: 3. Com he dit abans, hi ha hagut de tot, però el nivell el considero bo.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   El debat és més interactiu i engrescador per a la participació. És potser menys encorsetat que les bústies, fins i tot la del fòrum. El problema és que l'accés separat, el fet de tenir una guia de les intervencions o dels diferents fils a seguir més 'navegable' amb una doble finestra, per exemple, el fa més dependent de tota l'atenció.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Hauria de poder-se consultar el contingut del debat mentre estàs escrivint una rèpicl, o incloent un comentari. Per tal de poder fer referència millor a d'altres línies que hagin pogut aparèixer, o bé a d'altres alumnes i els seus comentaris. Estaria bé tenir un menú de navegació més intuitiu que et permetés fàcilment accedir a tots els fils i/o als comentaris que volguessis, veient-ne la interrelació.

Questionnaire #2

Preguntes sobre la nova eina Discussion Forum

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   4.

   Evitaria el fet de dir que 'No' per poder respondre o tancar el comentari.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   4 en els dos aspectes.

   Gràcies a la facilitat d'ús he pogut avançar molt en les tasques realitzades quan m'hi he pogut dedicar.

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   4 en els dos aspectes.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Avantatges totes, trobo que és un bon sistema, inconvenient un: Evitaria el fet de dir que 'No' per poder respondre o tancar el comentari.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   La resolució al desavantatge en el punt anterior.

Questionnaire #3

Preguntes sobre la nova eina Discussion Forum

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   4. Està molt ben pensada, tot i que trobo que es podria donar una organització més jeràrquica/esquemàtica per facilitar el seguiment del fils.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   3. Una organització més jeràrquica m'hauria ajudat a poder centrar-me més en determinats fils i participar més.

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   4. Crec que és una molt bona eina i molt ben pensada. En general, les persones de les TIC, ens agrada participar en noves idees i eines.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Els espais de l'aula són molt simples i bàsics. Aquest incorpora moltes funcionalitats que ajuden a organitzar treure'n un millor anàlisi de les contribucions.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Com ja he comentat, li donaria una organització més jeràrquica i cronològica per facilitar el seguiment dels fils.

Questionnaire #4

Preguntes sobre la nova eina Discussion Forum

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   (4) - L'eina me semble bona. En principi un tant complexe, però una vegada s'utilitza es senzilla i àgil.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   (3) - Crec que l'eina es molt dinàmica, amb lo qual facilita les intervencions, pot ser caldria una vista dels diàlegs en un fil un poc mes gràfica amb mes fàcil accés a les intervencions.

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   (4) L'eina ajuda a participar per la facilitat d'us i es mes dinàmica que el correu.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Com he dit abans es mes dinàmica i mes clara que el correu, en quan a problemes no he vist clar el tema de avaluar, crec que podria ser mes simple, i tampoc he vist com tancar un fil, si ja considerava que estava esgotat el tema.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Facilitar l'avaluació de les intervencions, pot ser clarificant millor, reduint les puntuacions a qualificacions del tipus: simple, mitjana, alta, etc. Mes participació de la figura de moderador, pot ser interessant en alguns casos marcar un poc línies de debat.

Questionnaire #5

Preguntes sobre la nova eina Discussion Forum

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   Pues en lineas generales la veo interesante, le pondria entre un 3 y un 4. La veo muy practica, añadiria alguna que otra cosa, pero bueno seguro que los compañeros que la estan desarrollando la mejoraran hasta conseguir un 5 de valoracion.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   No pienso que la herramienta me haya ayudado o facilitado para hacer aportaciones de calidad o cantidad.

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Idem de lo anterior.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Bueno si lo tengo que comparar con los debates del foro del aula, evidentemente esta herramienta le da mil vueltas. Las ventajas, casi todas, la creacion de carpetas e hilos, poder tener cada cosa en su sitio, las estadisticas, etc...

   No poder borrar mensajes (propios), lo encuentro una deficiencia.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Lo de borrar-modificar, podria ser interesante, ya que a veces te puedes equivocar.

Questionnaire #6

Preguntes sobre la nova eina Discussion Forum

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   Valoració 2. No m'ha agradat. Penso que és molt difícil seguir els fils de conversa. Possiblement sigui perquè no està preparat pel Firefox (almenys a mi alguns links no m'han funcionat com amb el IE). També penso que hi ha massa llibertat per escriure els missatges. Per exemple, alguns es poden marcar com a llegits, els altres si vols afegir alguna matisació no pots perquè has de dir que no per poder contestar,..

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Malgrat la dificultat de l'eina, aquest no ha estat motiu suficient per dificultar les meves exposicions. Una cosa molt bona són les estadístiques, això si que ha ajudat a veure l'estat de les meves intervencions.

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Valoració 3. Sincerament no crec que l'alta participació o la qualitat tingui una relació directe amb l'eina. En tot cas només es tracta d'una eina. Crec que ha estat més la motivació dels estudiants el que ha portat a que el debat hagi sortit d'una manera o altre.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Ja he comentat abans quins problemes he tingut. Possiblement remarcaria el fet que hauria de ser una eina senzilla i intuïtiva. I que no aporti masses opcions. Com més simple millor.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Una millora de l'aspecte visual ajudaria molt. Una idea seria agrupar intervencions per nivells i que aquests nivells es representessin gràficament amb agrupació de fitxes que amb els events del ratolí reaccionessin desplegant-se o agrupant-se segons el fil seleccionat.

Questionnaire #7

Preguntes sobre la nova eina Discussion Forum

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   Un 3. L'estructura hauria de ser una mica mes visual, podria ser mes en arbre. De vegades no saps quina aportació es mes recent.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Un 1. Sincerament crec que en res. Em sembla obvi que a nivell qualitatiu no te perquè influir, i a nivell quantitatiu tampoc tot i que bé, si et facilita molt la feina pot induir-te a participar.

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Entenc que igual que en mi mateixa, o com a màxim una mica mes. Un 2.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

Table 6.5: Main statistics extracted from the fifth experience using both discussion tools.

Statistics	**ST**	**DF**
Number of students	44	45
Number of threads	6	8
Total of posts	189	197
Mean number (posts/thread)	M=31.5 SD=2.7	M=24.6 SD=18.9
Mean number (posts/student)	M=4.3 SD=4.1	M=4.3 SD=2.9

Com ja he dit abans, penso que hauria de ser encara mes visual. Està clar que una avantatge respecte l'entorn habitual es que es millora nivell visual, però crec que ja que s'ha fet única i exclusivament per aquesta escomesa, doncs que hauria de ser molt mes visual, intuïtiva, etc... del que ja és.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Algun tipus d'exportació a un editor de text de tot un seguit d'aportacions. Molta de la gent que estudia a la uoc, ho fa perquè no te temps d'anar a la universitat presencial. Llavors molts tampoc tenen temps d'entrar cada dia. De vegades en un debat, estàs 3 dies sense entrar y tens 50 noves aportacions. Estaria ve poder imprimir o generar un document amb totes les aportacions filtrant per dates per exemple. D'aquesta manera pots llegir-les tranquil·lament en paper, que la pantalla sempre es mes incòmode.

Experience #5 - Spring term 2008

- Course: Methodology and Management of Computer Science Projects (MGPI)
- Assignment type and duration: Class discussion activity for 10 days
- Participation type: Required
- Discussion tools: UOC's *standard tool* (ST) and *Discussion Forum* (DF) version 0.6 (prototype)
- Potential participants: 44 (ST) + 45 (DF) graduate students

Statistic results

See Table 6.5[2].

Questionnaire results

Participation (optional): 13/45 (29%). Most relevant questionnaires' responses (original language):

Questionnaire #1

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

[2]Note this very high SD statistic in the posts/students mean. This was caused above all by two empty discussion threads, which were open by students just before closing the discussion. Without this outlier, SD=7.8. Please also note that, in this particular case, the number of ST's discussion threads were predefined and open by the lecturer only at the beginning of the discussion.

Nota: 3. Té moltes coses positives però que cal encara ajustar i millorar. Algun missatge del sistema, recordo no era l'adequat, no ajudava. Per exemple, en indicar (crec que era) bifurcar /estendre diàleg retornava un missatge dient que se n'havia de seleccionar primer, però resulta que no es podia seleccionar cap (no n'hi havia cap amb la casella. Per tant, en aquest cas, el missatge hauria de ser quelcom semblant a " ara no pots estendre un diàleg, per a fer-ho cal que...".

El tema de la separació de diàlegs s'hauria de mostrar d'una altra manera, potser en forma d'arbre. Però el problema més gran que li he vist és l'aparició dels missatge en ordre cronològic invers, la qual cosa és antinatural, a banda que molt molest.

Respecte les utilitats d'avaluació, assentiments i notes d'utilitat, doncs poden estar bé, però si s'obliga a la seva utilització, sinó la tendència és a no usar-ho i per tant no té cap utilitat pràctica.

La qüestió de les categories dels missatge està molt bé, encara que a vegades he tingut dubtes a quina categoria encaixaria millor el meu missatge (caldria definir miloor aquestes categories, o afegir-ne alguna més). De tota manera, la utilitat que pot tenir també em sembla dubtosa ja que la gent respon als missatge llegint els seus continguts, no llegint si la categoria és una petició o el que sigui.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Doncs en el meu cas particular ha influït bastant en la meva participació en quant a quantitat, no pas en quant a la qualitat de la participació. Tot i que considero que el nombre d'intervencions ha estat més que suficient, el fet de trobar-me amb un entorn amb el qual no estava familiaritzat i, per tant, no sentir-m'hi còmode (sobretot per l'ordenació cronològica inversa dels missatges) requeria per la meva part més temps de dedicació per, segurament i en un altre entorn més familiar, fer el mateix. I considerant la meva limitada disponibilitat de temps (ja ho he explicat abans), el fet d'haver-hi de dedicat temps a l'eina i no pas al debat en sí ha estat un handicap important.

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Nota 1 respecte l'alta participació en quant a quantitat. Crec que la nova eina no ha estat un factor determinat en l'alta participació en el debat. Crec que la participació en quant a quantitat no hagués estat més baixa si s'hagués utilitzat un espai col·laboratiu habitual. Inclús crec que la participació hagués estat superior.

   Nota 3 respecte a la qualitat de la participació. Potser aquí sí ha influït una miqueta més el fet de tenir classificats els missatges per diàlegs i categories: això pot representar una qualitat superior en quant a la classificació o ordenació dels missatge. Aquest fet pot revertir positivament en la qualitat dels missatges o intervencions. De tota manera, la qualitat dels continguts bàsicament depèn de les persones participants, i la influència de l'eina serà sempre molt menor que la decisió de cada participant en implicar-s'hi, en el debat, així com la seva capacitat o disponibilitat per a fer les intervencions de més o menys qualitat, nombroses, etc... Amb una altra eina col·laborativa habitual, la qualitat dels continguts no hagués estat significativament superior.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Per a poder fer útil les avaluacions (assentiments, notes) caldria obligar als participants a usar-les. Altrament les estadístiques associades poden no ser representatives de l'opinió del grup, per ser la mostra molt petita.

   Pensem que un projecte massa gran i complex té una probabilitat més alta de fracàs. Millor el dividim en projectes més petits, assequibles i amb resultats més immediats i visibles. Amb això vull dir, de manera metafòrica i comparativa, que si un fil té un nombre d'intervencions massa elevat, per al participant ja no es fa gens pràctic l'ús dels assentiments i notes. D'alguna manera s'hauria de fer que aquestes característiques es poguessin aplicar en agrupacions de missatges d'un ordre inferior (per ex. Diàlegs), i que es mostrés al participant un històric que indiqués quines i el nombre d'aquestes agrupacions (diàlegs) han estat ja avaluats.

El problema més greu és la ordenació cronològica inversa de les intervencions. També cal que, visualment, hi hagi una distinció mes intuïtiva entre els diàlegs. Manca la possibilitat d'adjuntar fitxers (almenys jo no l'he vista, aquesta funcionalitat), així com un enllaç directe al mail de cada participant, de tal manera que l'espai col·laboratiu seria més complert (síncron + asíncron, tot integrat).

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Ja ho he comentat al llarg del document.

Questionnaire #2

Preguntes sobre la nova eina Forum de Discussió

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   3. La recarrega de les pàgines fa molt lent processar els missatges quan la pàgina ja acumula uns quants i esdevé llarga. Crec que es podria millorar amb: poder veure les fotos dels companys, que els missatges estiguessin organitzats en un arbre desplegable, poder seleccionar grups de missatges per marcar-los com llegits, i finalment disposar d'una opció per contestar "NO esta d'acord" sense haver d'introduir un comentari, les opcions haurien de ser "SI"-"NO"-"Si, comentar"-"No, comentar".

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   3. L'estadística de participació la trobo molt interessant i m'ha motivat a fer més participacions, en quant a la qualitat no crec que l'eina la hagi afectat. No estic satisfet de la informació sobre la valoració quan he vist que participants amb una única participació han estat valorats amb A, jo amb la meva 1era participació vaig obtenir una B, i desprès d'haver afegit 3 intervencions la valoració ha canviat a una C+ (ja ho he comunicat al consultor, Santiago Codola, que m'ha donat una explicació).

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   1. Crec que l'eina no es clau en els aspectes mencionats ja que debats d'altres assignatures (d'aquest quadrimestre i d'anteriors) han tingut un nivell quantitatiu i qualitatiu equiparable, e inclús superior.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Les exposades en les respostes anteriors.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

Les exposades en les respostes anteriors, més la possibilitat treballar amb una opció ”Text enriquit” com es pot fer al crear nous missatges al campus virtual de la UOC.

Questionnaire #3

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   3

   No m'acaba d'agradar. En primer lloc, no crec que en debats d'aquest tipus siguin adequades les valoracions, ni que calgui manifestar-se a favor o en contra de les intervencions.

   En aquest sentit, prefereixo els espais de debat a l'aula, on si es creu convenient es matisa o es replica el comentari d'una altra persona. També he de dir que, particularment i en general, no crec gaire en els debats virtuals. En poques ocasions he trobat que siguin realment profitosos. Crec que els debats, normalment, necessiten la presencialitat.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   3

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   3

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Ja he comentat aquesta qüestió en respondre a la primera pregunta d'aquest apartat.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Bàsicament, no haver d'assentir o discrepar en gairebé totes les intervencions. Oferir l'opció de donar l'aportació per llegida.

Questionnaire #4

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   La meva valoració és de 4. Com aspecte negatiu trobo que la lectura de un fil amb molts de missatges és una mica difícil de seguir.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Un 4 en el grau quantitatiu. Potser ha fet que no veies sempre clar les respostes a les meves preguntes al no localitzar-les de forma ràpida.

Un 4 en el grau qualitatiu, per que puc incloure html i veure la pregunta a la que responc.

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Un 4, Caldria que avises amb emails quan algú ha contestat un missatge meu.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Com a millora Trobo que la visió global del fil és millor, per que en l'actual de la uoc, si algú contesta un fil molt antic costa de veure, i en el nou tenim un link als nous missatges a la part esquerra. Com a mancança trobo que l'autenticació no es del tot correcte per que un cop autenticat, puc enviar-me la url a un PC diferent i continuar posant missatges al fòrum.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Trobo a faltar que es puguin plegar i desplegar els fils de missatge, poder fer cerques amb cerca intel·ligent i també que guardi els filtres de cerca. Falta compatibilitat amb firefox, per que els links de nous missatges de la part esquerra no funcionen.

Questionnaire #5

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients): 2

   Parece una herramienta con futuro, pero ha dia de hoy todavia tiene muchos errores y genera continuas excepciones (entiendo que esta en fase beta o de pruebas). De cara a mejorarla, se deberia trabajar mucho mas la interfaz y hacerla mas intuitiva.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients): 3

   En cierto modo ha mejorado mi intervencion, ya que esta nueva herramienta mejora en parte la herramienta de foro de la UOC, y permite valorar las intervenciones y ver en que grado he intervenido.

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients): 4

   La herramienta a ayudado bastante pero pienso que gran parte del interes por la participacion de cada alumno esta en que esta sea obligatoria para aprobar la evaluación continua.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Permite a cada alumno valorar y medir el grado de interes de las aportaciones de sus compañeros. Muestra graficas y tablas con los detalles de la participación de cada alumno. En contra tiene que no esta integrado dentro del espacio de trabajo de la UOC, si no estras al foro no sabes si hay nuevas aportaciones, etc. Deberia ademas trabajarse mucho mas la usabilidad y que sea una herramienta mas intuitiva. Ya que pese a ser una herramienta nueva, no lo parece respecto a lo que estamos acostumbrados a utillizar en internet.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants. Aparte de lo que he comentado en el punto anterior, también se podria trabajar en que dentro de un debate apareciera mas claramente diferenciados las aportaciones de las respuestas o contestaciones a una aportación. Que se viese en forma de arboles o ramas.

Questionnaire #6

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Fòrum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   En un principi sorprèn que, tenint un fòrum de l'aula, o inclús carpetes d'altres assignatures que hi posi "Debat", calgui l'ús d'una nova eina. Ara bé, a mesura que t'hi incorpores hom s'adona que esta més ben dissenyada per al seu fi: el debat.

   No es tracta d'un espai on cliquem cada un dels missatges, sinó que podem seguir fils de forma independent, i llegir de forma molt més eficient les diferents opinions que s'hi fan.

   Valoració: 4

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Grau quantitatiu: 2

   Realment no crec que la nova eina sigui motivadora de més opinions. No veig el motiu pel qual, sigui en aquesta o per exemple al fòrum de l'aula, un estudiant hagi d'opinar més o menys.

   Grau qualitatiu: 4

   En aquest cas, i com he comentat a l'apartat anterior, la nova eina millora o facilita la qualitat de les intervencions, en el sentit que pots "motivar" el que intentes explicar al missatge, o per exemple, seguir les conformitats o no de la resta d'alumnes.

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Grau quantitatiu: 2

   Entenc que a la resta de companys de l'aula, la nova eina no els ha motivat més opinions. L'alta participació de l'aula, clarament segur que es deguda a que el debat es part de l'avaluació de la 1ª pac.

   Grau qualitatiu: 4

   Globalment és més ràpid i intuïtiu, i això afavoreix la qualitat de les intervencions en el sentit que es més ràpid llegir-les i seguir el debat, la qual cosa facilita que tots els alumnes puguem realitzar opinions més ben adreçades.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Millores: possibilitat de veure (tots alhora) i respondre els correus més ràpidament, disposar de la "Categoria" del missatge, poder crear diferents fils, posar nota (l'alumne) o unificar debats de diferents assignatures en una mateixa aplicació web.

   Mancances, problemes i/o inconvenients: mancança de FAQ, obligatorietat de respondre si no estàs d'acord amb una opinió (potser no estàs d'acord i no saps ben be com demostrar-ho), apartat d'estadístiques massa comprimit (si no et trobes a la part superior de la taula, costa entendre totes les columnes de les que es disposa).

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Malgrat la pestanya d'Ajuda, aquesta sembla un pèl escassa, tot i que existeixi un manual d'ús o una adreça de correu per a resoldre als dubtes.

   Potser la incorporació d'un conjunt de preguntes freqüents o FAQ, ajudaria en certs moments, així com reduir les columnes de les estadístiques a les més essencials, o anar-les reproduint a mesura que es va baixant a la taula en qüestió.

Questionnaire #7

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   La meva valoració es 1. Es molt poc ergonòmica, contínuament s'estan carregant els continguts per cada canvi (llegir,contestar, puntuar, etc.), esta obsoleta tecnològicament parlant. L'informe de participació esta molt bé, però el dia a dia es el debat.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   A nivell quantitatiu la meva valoració es de 4 (ha incidit molt) y a nivell qualitatiu 2 (ha incidit poc). Crec que el fet que sigui molt poc ergonòmica perjudica la quantitat de participació i que a nivell qualitatiu depèn més dels meus coneixements que de la pròpia eina.

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Crec que a nivell de participació el paper de la eina no ha incidit gaire (2) només podria afectar en el cas de veure a l'informe de valoració la teva classificació. A nivell de qualitat torno a comentar que es més una particularitat personal segons la preparació i experiència individual.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   En general les eines col·laboratives de la UOC tenen els mateixos problemes, informació dispersa, dificultat a l'hora de buscar la informació i problemes tècnics provocats per incompatibilitats amb els navegadors (encara que no son continuats).

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Poder buscar, adjuntar fitxers, un editor de HTML integrat, millorar els temps de carregar (fer servir més AJAX en la programació client), no queda molt clar la utilitat de les diferents categories. En general no trobo que sigui una eina intuïtiva com s'havia presentat i he de reconèixer que no vaig llegir de cap a peus el manual i segurament algunes de les meves dubtes es podrien solucionar amb una mica de estudi dels manuals.

Questionnaire #8

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   2. Tot i que el seu funcionament intern es molt bo, crec que encara es poc intuïtiu el seu us com a usuari normal.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   3. i 3. S'ha fet un poc dificil de llegir al no dispondre de paginació i no disposar d'un sistema d'iframes per a llegir les aportacions tan llargues fetes. El scroll s'ha fet un poc fatigat.

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   4. Pense que ha segut fàcil la tasca de publicar per part dels estudiants.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   M'agrada el tema del seguiment de puntuacions que podem fer i puntuar la qualitat d'un "post" però realment, avui en dia, em sent mes comode amb l'espai de debat de l'aula.

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Parlant des d'un punt de vista com a usuari que he sigut: - Edició del post; - Previsualitzar el post abans d'enviar; - Num. De caracters màxims; - El editor hauria d'incorporar-hi una eina per podar afegir taules, imatges, estil (http://www.fckeditor.net/demo); - Millora en l'estil (fonts, etc) de la pròpia interfície;

Questionnaire #9

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   Amb un 3. Aviam , l'eina funciona força bé però li calen millores. Hi hauria d'haver la possibilitat de donar més d'un missatge per llegit alhora o bé d'assentir les afirmacions de forma agrupada. És una mica engorrós que cada cop que cliquis el botó de donar per llegit o de respondre si es carregui la pàgina de nou i hagis de tornar a començar la lectura dels missatges. També m'ha passat que hi ha hagut algun missatge que no he pogut donar per llegit ja que estava tancat.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Ho valoraria amb un 3. M'ha ajudat a participar de forma quantitativa i qualitativa però li calen millores.

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   Crec que la gent en trobar-se amb una eina nova ha volgut "jugar" amb ella participant en els debats i contestant les afirmacions dels companys. També hi posaria un 3.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   Home, els debats de la UOC estan força bé. Com ja he comentat, aquesta eina necessita millores: possibilitat de donar per llegits uns quants missatges a l'hora, possibilitat de veure els mails dels companys per envia'ls-hi alguna consulta, caldria poder adjuntar fitxers, cercar missatges per data o per paraules,....

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   Caldria ordenar els missatges de forma més clara permetent que no et perdessis quan dónes un missatge per llegit, ja que es recarrega tota la pàgina i no saps on estaves.

Questionnaire #10

Preguntes sobre la nova eina Forum de Discussió

1. Com valores en general la nova eina Forum de Discussió per a la realització del debat. (valora 0 -5 i fes els comentaris que creguis convenients)

   4. És una innovació interessant ja que permet etiquetar i avaluar les intervencions.

2. Com creus que la nova eina ha incidit en el grau quantitatiu i qualitatiu de la teva participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   4 i 4. El fet de poder organitzar-se per fils de discussió és millor.

3. Quin paper creus que ha jugat la nova eina en l'alta participació assolida en el debat, com també en la qualitat de la participació. (valora 0 -5 per cada un dels dos aspectes i fes els comentaris que creguis convenients)

   4. Crec que ha estat elevada, tot i que en ser l'activitat obligatòria i puntuable per la PAC no sé si això hi influeix.

4. En comparació als espais de fòrum i debat tradicionals de l'aula, quines millores i avantatges creus que proporciona la nova eina per a la realització dels debats? Quines mancances, problemes i/o inconvenients creus que té respecte als espais de l'aula?

   - Possibilitat d'avaluar i etiquetar les intervencions.
   - Possiblitat suposo, pel professor, de treure estadístiques de la participació.

   Inconvenients:
   - Entorn fora del que és el campus virtual, manca d'integració.
   - Dificultat inicial d'aprendre l'entorn (usabilitat).

5. Quines millores introduiries a l'eina per millorar el suport que dóna als debats i als participants.

   - Possibilitat de tancar les banderetes d'una sola vegada i d'avaluar fils en conjunt. Fer-ho un a un és lent i es fa pesat.
   - Afegir eines de xat i videoconferència.
   - Possiblitat de penjar videos i col.laborar a temps real (web 2.0).

Table 6.6: Main statistics extracted from the sixth experience using both discussion tools.

Statistics	**ST**	**DF**
Number of students	141	139
Number of threads	60	10
Total of posts	140	242
Mean number (posts/thread)	M=2.3 SD=2.2	M=24.6 SD=4.1
Mean number (posts/student)	M=1.0 SD=4.9	M=1.7 SD=5.2

Experience #6 - Spring term 2008

- Course: Software Engineering (EP)
- Assignment type and duration: Two-class discussion activity for 14 days
- Participation type: Optional
- Discussion tools: UOC's *standard tool* (ST) and *Discussion Forum* (DF) version 0.6 (prototype)
- Potential participants: 141 (ST) + 139 (DF) undergraduate students, each group from two class rooms

Statistic results

See Table 6.6[3].

[3]Please note that, in this particular case, the number of DF's discussion threads were predefined and open by the lecturer only at the beginning of the discussion. Also note the high SD statistics of both tools in the posts/student mean. This may be interpreted by the optionality of the assignment and some students contributed a great deal while many others chose not to participate

APPENDIX B: Software development effort

The motivation of this appendix is to shed light on the software development effort made as the major contribution to this thesis. Indeed, among other small developments, six important software development projects of different size and purpose have been carried out over the last 5 years to provide full computational support to the the theoretical frameworks and conceptual models proposed in this thesis.

This brief overview shows the individual development effort made in a structured way in order to both ease its consultation and provide useful information on where to locate, download, and how to use this software. In overall, over 1,000 pages of technical documentation in the form of deliverables have been generated as well as over 200,000 lines of source code in different programming languages (e.g., Java, PHP, WSDL, and Python), among other development issues. This entire product can be reached via the Web links below (Web links as of April 2008).

Finally, the whole development has been supported by using *Open Source* (Webber, 2005) technologies only, such as Apache web server[4], Apache Tomcat[5], Apache Axis[6], PHP[7], Java[8], Python[9], Zope[10], Plone[11], PEAR[12], ZSI[13], and PostgreSQL[14], and the entire software is to be registered to SourceForge[15] community shortly .

General Purpose Library

See Table 6.7.

[4] The Apache Web Server project is an open-source HTTP server for modern operating systems including UNIX and Windows NT. It is found at: http://httpd.apache.org/ (Web site as of April 2008)

[5] The Apache Tomcat is an Open Source JSP and Servlet Container from the Apache Foundation project is found at: http://tomcat.apache.org/ (Web site as of April 2008)

[6] The Apache Axis project develops a Java platform for creating and deploying web services applications. It is found at: http://ws.apache.org/axis/ (Web site as of April 2008)

[7] PHP is a server-side HTML embedded scripting language. It is found at: http://www.php.net/ (Web site as of April 2008)

[8] Java is a programming language originally developed by Sun Microsystems and released in 1995. It is found at: http://java.sun.com/ (Web site as of April 2008)

[9] Python is an interpreted, interactive, object-oriented, extensible programming language. It is found at: http://www.python.org (Web site as of April 2008)

[10] Zope is an open source application server for building content management systems, intranets, portals, and custom applications. It is found at: http://plone.org/ (Web site as of April 2008)

[11] Plone is a Open Source Content Management System (CMS). It is found at: http://www.zope.org/ (Web site as of April 2008)

[12] The PHP Extension and Application Repository (PEAR) is a framework and distribution system for reusable PHP components. It is found at: http://pear.php.net/ (Web sire as of April, 2008)

[13] The Zolera SOAP Infrastructure (ZSI), is a Python package that provides an implementation of SOAP messaging. It is found at: pywebsvcs.sourceforge.net/zsi.html (Web page as of April 2008)

[14] PostgreSQL is a powerful, open source relational database system. It is found at: http://www.postgresql.org/ (Web site as of April 2008)

[15] SourceForge is the world's major open source software development web site, located at: http://sourceforge.net/ (Web site as of April 2008)

Table 6.7: **GPL project:** Description and development effort.

Project description	The General Purpose Library is made up of components of greatly generic use that creates the skeleton for the construction of complex systems requiring the management of the users interacting with the system and optimization of the system's resources.
Start date and current version	September 15, 2002; v1.0 June 20, 2003
Development docs	http://clpl.uoc.edu/docs/GPLdevelopment.zip
Source code	http://clpl.uoc.edu/src/gpl-java.zip
Binaries	
Web site / API	http://cv.uoc.edu/šcaballe/tfc/api
Size and development effort	210 pp UML-based specification and design 8,000 Java code lines (approx)

Collaborative Learning Purpose Library

See Table 6.8.

EventExtractor

See Table 6.9.

UOCLogsProcessing

See Table 6.10.

Discussion Forum

See Table 6.11.

Communities of Learning Practice Environment

See Table 6.12.

Table 6.8: **CLPL project:** Description, downloads, and development effort.

Project description	The Collaborative Learning Purpose Library (CLPL) creates the skeleton for the construction of specific Computer-Supported Collaborative Learning (CSCL) applications. The component development is based on the Generic Programming paradigm permitting their complete and effective reutilization within the CSCL domain. This library is made up of five components and represents a particularisation from the Generic Purpose Library (GPL) which is indispensable to be able to use the CLPL.
Start date and current version	October 1, 2003; v1.1 January 31, 2008
Development docs	http://clpl.uoc.edu/docs/CLPLdevelopment.zip
Source code	http://clpl.uoc.edu/src/clpl-java.zip (Java version) http://clpl.uoc.edu/src/clpl-wsdl.zip (web-service version)
Binaries	http://clpl.uoc.edu/bin/clpl.zip
Web site / API	http://clpl.uoc.edu/ http://clpl.uoc.edu/downloads/CLPL1.1released.zip
Size and development effort	430 pp UML-based specification and design 20,000 Java code lines (approx) 156,000 WSDL code lines (approx)

Table 6.9: **EventExtractor project:** Description, downloads, and development effort.

Project description	Log file processor for the BSCW application. This application runs offline on the same machine as the BSCW server and uses the daily log files generated by the BSCW server as input so as to: (i) identify the event boundaries inside the log file, (ii) map specific information contained in these events about users, objects, sessions, etc. to typed data structures, and (iii) store these data structures in a persistent support. Both sequential and grid versions are available
Start date and current version	December 11, 2004; v4,0 August 31, 2006
Development docs	http://clpl.uoc.edu/src/EventExtractor.zip
Source code	http://clpl.uoc.edu/src/EventExtractor.zip
Binaries	http://clpl.uoc.edu/src/EventExtractor.zip
Web site / API	http://clpl.uoc.edu/src/EventExtractor.zip
Size and development effort	2,000 Java code lines (approx)

Table 6.10: **UOCLogsProcessing project:** Description, downloads, and development effort.

Project description	Log file processor for the UOC virtual campus. UOCLogsProcessing application runs offline on the same machine as the logging application server. It uses, as an input, the daily log files of the virtual campus obtained as a result of merging all web servers' log files. The following process is run: (i) identify the log entries boundaries and extract the fields that make up each entry, (ii) capture the specific information contained in the fields about users, time, sessions, areas, etc., (iii) infer the missing information, (iv) map the information obtained to typed data structures, and (v) store these data structures in a persistent support.
Start date and current version	March 30, 2006; v2.0 May 31, 2006
Development docs	http://clpl.uoc.edu/src/UOCLogsExtractor.zip
Source code	http://clpl.uoc.edu/src/UOCLogsExtractor.zip
Binaries	http://clpl.uoc.edu/src/UOCLogsExtractor.zip
Web site / API	http://clpl.uoc.edu/src/UOCLogsExtractor.zip
Size and development effort	1,200 Java code lines (approx)

Table 6.11: **DF project:** Description, downloads, and development effort.

Project description	The Discussion Forum is a web-based knowledge-based structured collaborative learning system aiming at supportting and fostering collaborative knowledge building by means of asynchronous discussion process. It provides essential functional support for discussion threads, dialogs and contributions. It extracts relevant knowledge about group activity in order to provide learners and tutors with efficient awareness, feedback, and monitoring as regards learners' performance and collaboration.
Start date and current version	January, 15, 2007; v0.8 January 31, 2008
Development docs	http://clpl.uoc.edu/docs/DiscussionForumSpecification.pdf http://clpl.uoc.edu/docs/DiscussionForumDesign.pdf
Source code	http://clpl.uoc.edu/src/df.zip
Binaries	http://clpl.uoc.edu/src/df.zip
Web site / API	http://clpl.uoc.edu/df/
Size and development effort	200 pp UML-based specification and design 8,000 PHP code lines (approx) 90,000 Java code lines (approx) 1,000 configuration lines (approx)

Table 6.12: **CoLPE project:** Description, downloads, and development effort.

Project description	Communities of Learning Practice Environment is a technology-enhanced learning tool to support asynchronous discussion in communities of formal learning practice. It provides essential functional support for democratic groupware. It extracts relevant knowledge about group activity in order to provide learners and tutors with efficient awareness, feedback, and monitoring as regards learners' performance and collaboration.
Start date and current version	September, 1, 2007; v0.6 February 23, 2008
Development docs	http://clpl.uoc.edu/docs/CoLPESpecification.pdf http://clpl.uoc.edu/docs/CoLPEDesign.pdf
Source code	to be completed
Binaries	to be completed
Web site / API	http://clpl.uoc.edu:8080/colpetest
Size and development effort	220 pp UML-based specification and design Unknown Zope/Plone-based, Python and Java code effort