UNIVERSITAT DE BARCELONA

DEPARTAMENT DE BIBLIOTECONOMIA I DOCUMENTACIÓ

Doctorat en Informació i documentació en l'era digital

Programa interuniversitari de doctorat 2006-2008 (UB-UAB)

Towards a personalised virtual library: indications from navigational and personal information behaviour of e-learning students

Tesi doctoral que presenta **Núria Ferran i Ferrer** per optar al títol de doctor per la Universitat de Barcelona

Directors: **Dr. Julià Minguillón Alfonso** i **Dr. Mario Pérez-Montoro Gutiérrez**

Barcelona, gener de 2010

A la Queralt i al Jaume

Contents

Acknowledgements

I am truly grateful to my PhD dissertation supervisors Professor Julià Minguillón and Professor Mario Pérez-Montoro for their wisdom and for supporting my research ideas and initiatives. I would like to thank the research and administrative staff from IN3, Universitat Oberta de Catalunya and the Department of Biblioteconomia i Documentació de la Universitat de Barcelona, who shared their knowledge with me and helped me in my development as a researcher.

I also appreciate the financial support for the field research work provided by UOC-IN3 as a result of coordinating the European research project OLCOS (EAC/23/05 AT 002) and the project PERSONAL(ONTO) ref. TIN2006-15107-C01.

Many thanks to the anonymous reviewers of the four papers presented in this thesis dissertation whose comments have substantially improved this research work.

I am very grateful to Professor Sheila Webber who organized and hosted my research stay at the Information Studies Department at Sheffield University which was a very fruitful and pleasant period of my PhD degree and hopefully our work together will continue through the Information Literacy Research Centre (ILRC). I deeply appreciate her help and guidance during this stay. I am also very thankful to Professor Nigel Ford from the Information Studies Department and Professor Bill Johnston from the University of Strathclyde.

My special thanks are for Dr. Jordi López Sillas from the Universitat Autònoma de Barcelona for his comments and suggestions; Muriel Garreta, Dr. Enric Mor and Mireia Pascual for their invaluable collaboration; and to Laura Rojas for her support during the field work.

My gratitude also extends to several people who commented on parts of this PhD dissertation, namely the participants of the Human Computer Interaction International Conference 2009, the different editions of the Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables (SPDECE), the members of the European research projects COINE and OLCOS, Prof. Sirje Virkus from Tallin University, Gill Needham from the Open University Library, Professors Kari, Savolainen and Sormunen from Tampere University, Monika Krakovska from

Krakow University and Sandra Schaffer of EduMedia from Salzburg Research. I would also like to thank Alice Keefer from the Universitat de Barcelona for all her enthusiasm throughout the course of this study.

I would like to acknowledge the interest shown in my research and the support from my colleagues in the Information and Communication Science Studies of the Universitat Oberta de Catalunya and a special thanks to Sandra Vilajoana, Dr. Anna Clua and the Director, Dr. Lluís Pastor and to my colleagues at the Facultat de Comunicació de la Universitat Autònoma de Barcelona; and last but not least, to my fellows of the research group NET2LEARN.

I am also indebted to all those who supported my work from the UOC digital library; to Kathryn for her English assistance and to the SAIC Summer Seminars for their useful insights and suggestions.

I thank my friends Carme, Isabel, Francesc, Natasha, Yulia, Noemí, Sam, Eva, Iolanda and Glòria for sharing my preoccupations and for the enjoyable time during my PhD degree.

I express my deepest gratitude to my family, especially mare, pare, Joan, Montse, Joana, Maria (nephew), Marta, Maria (cousin), Rosa, Miquels i Eu for their encouragement, help and love.

Last but not least, to Queralt who has given me the right balance between the focus and the distraction and to Jaume for interchanging ideas with me, for his sharp critical thinking, his brilliant comments, creativity, experience, patience, enthusiasm and, above all, for his unconditional love.

Overview

To better understand information behaviour and how to make information delivery systems more usable, this dissertation endeavours to investigate how people search, manage and use information resources. More precisely, this research focuses on the information behaviour of e-learning students and how such behaviour can be used to improve the design of personalised information delivery systems in order to aid in the fulfilment of their learning goals. One direction for improvement is personalisation, which is strictly linked to user satisfaction and provides efficient and effective support towards the process of seeking and using information. Non-authoritative metadata automatically generated from the interaction of users with resources can be a practical solution to enhance personalised information delivery systems.

In addition, the context of e-learning students is not exclusively academic, as they also develop their information behaviour in other contexts such as in the workplace and daily life. In the case of the Universitat Oberta de Catalunya, there is the opportunity to study the full context of the information behaviour of individuals focusing on the e-learning environment, but also including the rest of the spheres where they develop their activities. Furthermore, this PhD dissertation studies the knowledge, attitude and skills that allow individuals to informationally behave in a particular manner. The acquisition of information literacy may take place in one context and be later transferred to others. In terms of added value, e-learning universities should provide their students with the adequate level of information literacy in order to help them adopt the appropriate information behaviour for their information age.

Overall, this PhD dissertation provides some indications towards a personalised design of information delivery systems for e-learning environments and for the generation of learning plans that are more adapted to the requirements of lifelong learners. In this sense, this dissertation brings about a definition of the elements for designing personalised virtual libraries and proposes some improvements in the description of the learning object metadata, based on the usage of educational resources.

It also provides theory and experimental evidence on how to model information behaviour patterns taking into account all the informational contexts in which a user can develop his or her information behaviour.

Four papers included in this thesis address different research questions related to the outlined topic.

Paper I

FERRAN, N.; MOR, E.; MINGUILLÓN, J. "Towards personalization in digital libraries through ontologies", *Library Management Journal*, 2005, vol. 25, no. 4/5, p. 206-217.

This paper is based on the basic premise that the efforts for finding an information resource in a digital library can be stored in a structured way and then shared by future users with similar necessities. Digital libraries can offer personalised services which can adapt content and its presentation to different user profiles, thus bringing about personalisation which is one of the key factors for user satisfaction. This article defines the elements for building a personalised information delivery system integrated in an e-learning environment. The description of the relationships between those elements by means of an ontology is provided. More precisely, the methodology used for designing ontologies is applied in order to define the requirements of a personalised information delivery system. Two elements determine the functionalities of this desired personalization system: firstly, user profiles, including the navigational history and user preferences; and secondly, the information collected from all the navigational behaviour developed by digital library users. Both elements are analysed in the article and described in depth.

Paper II

FERRAN, N.; CASADESÚS, J., KRAKOWSKA, M.; MINGUILLÓN, J. "Enriching e-learning metadata through digital library usage analysis", *The Electronic Library*, 2007, vol. 25, no. 2, p. 148-165.

This article studies the interaction between user profiles and information resources. It presents several experiments carried out in a virtual learning environment in order to propose an evaluation framework for analysing learning object usage, with the aim of extracting useful information for improving the quality of the metadata used to describe such learning objects, but also for personalisation purposes, including user models and adaptive itineraries. This article reveals some interesting facts about the real usage of

the learning resources which can be added as new metadata, overcoming information overload by means of proper recommendations, quality assessment and improving the learning process itself once it is described through formal descriptions. A selection of IEEE LOM (Learning Object Metadata) fields is provided to enrich the description of learning objects, although automatically in some cases, from the use of the information resources carried out by the users of an academic community.

Papers III and IV

FERRAN, N.; GUERRERO-ROLDÁN, A.; MOR, E.; MINGUILLÓN, J. User Centered Design of a Learning Object Repository", M. Kurosu (Ed.): *Human Centered Design, HCII 2009, Lecture Notes in Computer Science*, no. 5619, p. 679–688.

FERRAN, N.; PÉREZ-MONTORO, M. La Gestión de la Información Personal en usuarios intensivos de las TIC, *El Profesional de la Información*, 2009, vol. 18, no. 4, p. 365-373.

The principal objective of these papers is to explain the variables involved in information behaviour. The analysis of information behaviour actions related to personal information management is crucial for designing virtual environments that support and promote the activities that an individual carries out to acquire, avoid, create, store, organise, maintain, retrieve, use and distribute the necessary information in the completion of tasks in order to fulfil different roles and responsibilities: e.g. being a parent, student, worker or member of a community. The virtual campus becomes a common space where students acquire, use and develop information competencies from any of the three possible contexts (workplace, daily life and academic). It is this transfer of competencies that has a significant influence on the way a student tackles the obvious need to search and use all the information that is generated during the learning process.

The focus of **paper III** is on learning object repositories but in particularly, taking into account that the new European Higher Education Area (EHEA) promotes the design of learner-centred processes that focuses on the acquisition and development of competencies rather than on the consumption of contents. This is the reason why it is suggested in this paper that learning objects are much more than just systems for providing access, reusability and preservation of contents. What is more, they should also be fully integrated into the learning process. In this sense, a qualitative

methodology approach is presented as part of an information behaviour analysis as the background to integrating learning object repositories in virtual learning environments following the user-centred design requirements.

In **paper IV**, the qualitative analysis goes one step further and provides the identification of different user profiles as the first stage to design a user-centred personalized information delivery system. Four user patterns of information behaviour were identified through combining two studied variables (cognitive approach and attitude): exhaustive, passive, reactive and proactive. These patterns show the differences in information behaviour for each context (daily life, workplace and academic). The identification of the user profiles is the essential first step in a user centred service design that addresses the specific intervention suitable for each user type, taking into account the tool and process requirements needed by them to develop efficient standards of information behaviour.

Although all the articles are interrelated and complementary, each of them represents an independent unit of research with its own motivation, literature review and research questions. These papers can be found in the Annex section of this PhD dissertation. They are presented together with a "General Introduction", in Chapter 1, which provides the literature review, the purpose and significance of this PhD dissertation, the work hypothesis, research questions and the methodology used in the different studies. Chapter 2, "Main Results", offers a brief statement with the principal outcomes achieved in the study, a summary of the discussions and some indications for a more improved designing of information delivery systems and for supporting the creation of learning plans. Finally, these deliberations lead to Chapter 3 where the conclusions are laid out.

1. General introduction

This chapter presents the main concepts of this PhD dissertation through a literature review. Subsequently, the purpose and significance of the research work is outlined as well as the hypothesis, research questions and the methodological approach to study the shift towards the personalisation of a virtual library, through analysing web navigational and informational behaviour of e-learning students in their academic, workplace and daily life contexts.

1.1. Literature review

The widely accepted definition of **information behaviour** is the one given by Wilson (2000), that embraces human actions in relation to information sources and channels, including both passive and active information seeking and the use of it (Fisher and Julien, 2009) or its transfer (Wilson, 1999), as well as the behaviour aimed at avoiding information (Case, 2002).

Research into human behaviour in relation to the search and use of information has evolved over time. At the start of the twentieth century, studies began with library users, focusing primarily on determining which services they used, how often, what information they used, and all those aspects related with the sociological and demographical characteristics of the group to which they belonged, without concluding what use the individuals attributed to the library as a source of information (Bawden, 2006). This traditional perspective was concerned with the process of information seeking, as it was in these activities that the user came into contact with information system (e.g. the library) or used specific information (González Teruel, 2005).

Subsequent studies began in the middle of the last century into how people used information in relation to their work. However, these and recent studies focused on how information systems were used and did not look into the information behaviour and information needs of users (Wilson, 2000). Not until the 1970s did research start focusing on the person as an information seeker, creator and user (Case, 2002). In

particular, this change of direction in research focused on the user instead of the system occurred with the studies carried out by Tom Wilson (1981), the work about information needs for everyday life (Dervin, 1976; Savolainen, 1995) or the work of Carol Kuhlthau (1991).

Information behaviour is a field of study that comprises three constituent elements: the need for information; the process for the search of information and the use of information (Wilson, 1999). Although these elements have been studied to a different extent and the most studied and modelled are the ones concerning search needs and the search process; in contrast, the use of information is the aspect that has received much less attention and is much less defined (Wilson, 1999; Case, 2002). The following is a description of the different studies which focus on information needs, the search process and the use of information:

a) The need for information was one the main reasons why information seeking processes started to achieve special significance, however, there are other informational activities unrelated to information needs, such as entertainment, company needs, etc. (Allen, 1996). Information behaviour models that develop the information seeking part and pay no attention to its use, situate information needs as the context in which users use information systems. The need for information becomes visible through the information search process, and it is through the result of the search process that, these needs are met or not (Allen, 1996). Information needs are related to the use of information in the sense that information is needed for a specific use. When the needs are studied, the aim of information seeking is discovered but it does not go further to know how to apply the information that is discovered (Barlett and Toms, 2005).

b) Information seeking behaviour allows the cognitive and intangible aspects of the concept of the need for information to be put aside and then the aspects that are desired to be observed can be clarified. According to Wilson (1981), information seeking behaviour would become the behaviour of an individual when he or she needs information to reach an aim. Another point of view is the Sense-Making theory in which information seeking is seen as a dynamic sense allocation process where the user embraces cognitive behaviour at an individual perception level and communicative

behaviour at a social context level (Ellis, 2005). Information seeking behaviour has also been studied as a construction process similar to learning, based on the factor of uncertainty which motivates the individual who searches information (and also the student) to start the seeking and learning process (Kuhlthau,1997). And last but not least, in more recent times, the seeking process has been defined as a complex activity related to information and communication that requires access to different information sources to deal with personal, social and professional problems (Spink and Cole, 2004).

Information seeking behaviour has at its disposal numerous models that help to understand it; they have been revised in a chapter of the *Annual Review of Information Science and Technology* (Pettigrew, Fidel and Bruce, 2001). In particular, the models with the greatest impact are those by Wilson (1981), Krikelas (1983), Dervin (1983), Kulthau (1991) and Jones (2007).

c) Information use is that part of Information Science that examines what people do with information once they have obtained it. It is often linked to information needs, in the sense that information is needed in order to be further used, so it could be seen as the factor that drives other information activities as it represents the reason why information is needed and sought (Barlett and Toms, 2005). Another point of view to be examined is what happens with the information that is obtained and how it is applied to specific tasks or to meet objectives (Allen, 1996).

Very little study has been conducted into the actions related to the use of information. To illustrate this fact, we ran a keyword search[1] on the Library and Information Science Abstracts (LISA) database of Cambridge Scientific Abstracts, in the last ten years (1999-2009), which provided for us some quantitative evidence on this fact. With the "information need*" query, 974 peer-reviewed articles were found; with 1,314 on "information seeking"; while 329 on "information use*" and only 14 on "information usage*". It is paradoxical that despite the relevance of understanding the user in the design of information systems to meet demands, the use of information is one of the least studied areas of research (Sanz Casado, 1994; Tuominen 1996 quoted in Kari, 2007). At the same time, this demonstrates that there is a serious weakness in

[1]The keyword search in LISA database retrieves matching results coming from the title, abstract, descriptors and identifier fields.

Information Science as information is not just acquired for its own sake (Dervin, 1999) but rather to be used.

Nevertheless, there have been several approaches to study **information use**. There are few theories that explain this process and there are no clear or accepted models compared to the seeking process as it has been above mentioned. However, we can talk about different approaches: Information Routines (Mintzberg *et al.*, 1976; quoted in Choo, 2002), the Taxonomy of the Uses of Information by Taylor (1991), the Three Steps Model by Saracevic and Kantor (1997), the theorisation by Choo on knowledge creation (2002) and the conceptualizing of information outputs (Kari, 2007).

From the management and business sphere, Mintzberg *et al.* (1976) distinguishes three information routines: "exploration", related to information scanning and revision phases, both if they have been obtained actively or passively; the "research" routine, which collects and analyses specific information required during the activities of diagnosis, seeking and evaluation options; and, finally, the "dissemination" routine, when information on the decision-making process is made known.

Another proposal for modelling the use of information is Taylor's taxonomy (Taylor, 1991) which is the expression of the perception of information needs rather than a set of uses; consequently, the information uses that it detects include: factual (looking for a specific piece of data), confirmation (checking information) and instrumental use (what to do and how to use something).

The Saracevic and Kantor model (1997) keeps to a very general level. It describes the acquisition, cognitive and application steps, without a precise specification of which ones form part of the seeking behaviour and which ones relate to its use. Furthermore, it does not explain which of the different usage contexts the information can be placed.

With regard to Choo's conceptual proposal (2002), it understands the use of information as the creation and application of knowledge to decision-making processes: the "Identification" phase, where a problem is recognised and diagnosed; the "Development" phase, where solutions are sought; and the "Selection" phase, where inappropriate alternatives are eliminated, the facts are judged and analysed and a decision is made.

Finally, Kari (2007) proposes organising the typology of the generic uses of information into two groups: the social, which includes uses related to communication and therefore the demonstration of knowledge; and personal, which refers to doing (physical uses) and thinking (the mental process for interpreting the information).

The most widely performed research to date has been focused on the uses of individuals with e-mail (Balter, 2000; Bellotti, Ducheneaut, Howard, Neuwirth, and Smith, 2002; Bellotti, Ducheneaut, Howard, and Smith, 2003; Bellotti and Smith, 2000; Ducheneaut and Bellotti, 2001; Gwizdka, 2000, 2002a, 2002b; Mackay, 1988; Whittaker and Sidner, 1996; Wilson, 2002; quoted in Jones, 2007). Numerous studies have also been carried out into the use of the Web, mainly in activities such as bookmark management or browsing history (Abrams, Baecker, and Chignell, 1998; Byrne, John, Wehrle, and Crow, 1999; Catledge and Pitkow, 1995; Tauscher and Greenberg, 1997a, 1997b; quoted in Jones, 2007) and also the organisation and management of paper and electronic documents (Carroll, 1982; Case, 1986; Malone, 1983; Whittaker and Hirschberg, 2001; quoted in Jones, 2007). There are no studies that analyse informational needs in a virtual learning environment.

Another theory with a different approach to study information use behaviour comes from the literature on **information literacy**. Information behaviour and information literacy are linked because, in order to become information literate, one has to acquire an appropriate information behaviour (Boon *et al.*, 2007). To help to theorise and specify information use behaviour, literature on information competences can be useful as a foundation, since information literacy contemplates both seeking and use processes. According to the ICT Literacy Panel, information behaviour can be displayed on the basis of different actions (2002):

- Access to information: gathering and retrieving information. This involves what an individual wants to access, the reason why and the locating and the retrieving of information.
- Information treatment: applying existing organisation or classification systems.
- Integration: interpreting and representing information. This involves

summarising, comparing or proving.

- Evaluation: generating opinions on the quality, relevance, usefulness and efficiency of information.
- Creation: generating information, adapting it, applying it, designing it, inventing it, etc. This involves applying information critically to solve a specific question, by creating new information or recreating existing information.

A sixth action could be included to information behaviour in order to complement the aspects related to the use of information:

- Communication: as SCONUL points out in the Seven Skills Pillars (2007), it refers to communicate effectively using appropriate media, according to specific aims or recipients and to understand issues of copyright and plagiarism. To this end, the most suitable styles, types, tools and channels for disseminating information or communicating should be used.

However, while some definitions of information competences (UNESCO, 2003) have an individual perspective in which the person identifies, evaluates and uses information without having contact with other people, there are other definitions where the concepts have a more communicative connotation (Bruce, 1997), which deal with the influence of communication on the identification, evaluation and use of information. In this last trend, the main topic is usually not information competences but rather personal knowledge management (Schreiber and Harbo, 2004). We understand that both topics are opportune for the academic and workplace context and even daily life. Everyone cooperates with other people or is influenced by others, both in the learning process and when it comes to creating knowledge (Schreiber and Harbo, 2004).

The concepts of information behaviour and competences do not appear explicitly together in academic and professional literature. There appear to be two strands of study; on the one hand, studies about information seeking practices, often from the point of view of training in the use of libraries and information systems (Lloyd, 2006); on the other hand, studies into knowledge and skills. However, there is a need to study them jointly, as information competences are intrinsically associated with information

practices and critical thought in the environment provided by ICT (Bruce, 2002). Only a few studies, although to be taken seriously, have discussed both concepts together, but issues such as everyday information behaviour have not been explored (Boon *et al.*, 2007).

Often, when one speaks of digital training, information competences are put at the same level as technological competences. Nevertheless, there are two sets of different skills that together define digital training. On the one hand, the main aim of technological competence is to make people independent in the use of different tools, primarily those of ICT, and therefore to be able to use them efficiently and effectively, in order to achieve specific aims. ICT skills may include those basic skills such as the use of the keyboard, mouse or printer; the use of standard software which includes the word processor, spreadsheets and databases, among others; and network applications which include electronic mail and the Internet. On the other hand, and distinct from IT skills, there are information handling skills which include making a proper use of information sources, evaluation criteria, navigation methods, manipulation techniques, and presentation issues (Corrall, 1998).

This kind of distinction is supported by other authors, who also challenge the tendency to equate computers with information and hence to misunderstand computer literacy for information literacy (SCONUL, 1997, Eisenberg, Lowe and Spitzer, 2004). “This is a dangerous myth, as it assumes that information is only that which is storable and manipulable in a computer” (Taylor, 1991).

Information literacy is therefore a much broader concept which refers to the adoption of appropriate behaviour to obtain information, through whatever channel or medium, and which is well fitted to information needs, together with a critical awareness of the importance of the wise and ethical use of information in a society (Johnston and Webber, 2003). Other definitions are more focused on particular actions involved with information behaviour and refer to information literacy as the ability to recognise an information need and the ability to identify, find, evaluate, organise, communicate and use the information effectively, both for solving problems and for lifelong learning” (AASL, 1998).

Information competence, in the education sector, is analysed through taking into account as a premise that such a competence facilitates the learning process on the basis of the correct use of information. In the academic sphere, information competence is considered to be "the use of information in all its dimensions: access, analysis, interpretation, evaluation, production, etc." (Picardo, 2002). In fact, with these competencies, information is a previous step to knowledge (Toledano, 2008). By contrast, in the professional environment, the concept is related to training that should enable workers to access and use the information involved in the execution and performance of their daily tasks (Bruce, 1997).

Research projects recently conducted suggest the existence of two sub-categories of information competences; those specific to the environment where they are carried out and transverse ones (Bundy, 2004). The specific competences are contextualised, depending on the environment where they are carried out (Lloyd, 2006), whereas the transverse ones can be transferred between environments. So, for example, in the academic sphere, a great deal of emphasis is placed on the competences related to the information seeking process; by contrast, in the professional sphere, the competences that can be put into practice are those related to the access and use of information, while, finally, in the daily life environment, the competences related to communication through socialisation tools are most developed.

As the analysis of the competences that are transferable from one sphere to another is one of the least studied aspects in terms of the competences developed when searching for and using information (Lloyd, 2006), it would be interesting to determine whether and to what extent there is a transfer of information competences. According to some authors, this question is not conclusive (Misko's, 1998, Anderson, *et al.*, 1996; quoted in Lloyd, 2003). This is the main idea which this thesis articulates.

1. 2. Purpose and significance of the study

Today, **information behaviour** is intensely studied in a number of spheres, such as the professional context among doctors, researchers, IT specialists, the fire service, etc. (Wilkinson, 2002; Leckie, 2005); in private life, in relation to leisure, health, domestic finances, etc. (Case, 2002; Savolainen, 2005); or in the academic sphere, among first-cycle students, postgraduate students, PhD students, scholars, etc. (Fulcher and Lock, 1999; Talja, 2005; Thósteinsdóttir, 2005; Nowé, *et al.* 2008), There is a recognised need to develop usable designs according to the context of the information consumers in order to meet their information needs (González Teruel, 2005; Hepworth, 2007).

From a social and economic perspective, the study of human behaviour in terms of information sources and channels and its further use is important, since we are immersed in the informational society in which the generation, processing and transmission of information becomes the fundamental sources of productivity and power due to the technological conditions of this historic period (Castells, 1999). And from an individual point of view, the analysis of information behaviour actions related to personal information management is crucial in the design of virtual environments (Severance et al., 2008) that can support and promote personal information behaviour; referring to the activities that a person carries out to acquire, avoid, create, store, organise, maintain, retrieve, use and distribute the necessary information to complete tasks in order to fulfil different roles and responsibilities: e.g. being a parent, student, worker or member of a community (Jones, 2007).

Hence, there is a challenge in information delivery systems to maximize information seeking and searching by means of **personalisation**, in order to increase their effectiveness and efficiency. Personalisation is also linked to user satisfaction (Riecken, 2000), provided that it is properly introduced (Nielsen, 1998), as has been shown in several studies in areas such as e-commerce (Kasanoff, 2001; Kleinberg and Sandler, 2008), business-to-business (Colkin, 2001, Dyché, 2002), e-learning virtual environments (Mor and Minguillón, 2004; Dongming and Huaiqing, 2006) and specifically into library initiatives (Sinha and Swearingen, 2001; Neuhold *et al.*, 2002). However, in order to implement personalised information delivery systems, they must

be addressed from a user centred design and, therefore, they should address aspects such as behavioural aspects (Marchionini, 1995; Smith *et al*., 1997; stated in Cunliffe, 2000; and Schonberg *et al.*, 2000).

It is worth remarking that "conceptually, personalization is easy to understand, but in practice it is difficult to implement" (Law *et al*., 2006). The concept of web personalisation can be defined as a "way to meet web users' requirements in order to provide precise information that best suit the user's need and to provide useful recommendations" (Nayak and Jain, 2008). One of the most successful implementations for providing recommendations in systems with a large number of users is collaborative filtering (Herlocker, *et al.*, 2004). This technique for personalisation is based on an automatic system which collects information on user actions (explicit, such as voting or answering a question related to the performed action, or implicit, such as noticing which offered links are visited and which are not) and determines the relative importance of each piece of content by weighing up all the collected information among the large amount of users. Therefore, according to some criteria, this kind of personalised system uses observable behavioural patterns, user preferences from indirect sources as well as methods for inferring user behaviour. It takes into consideration the behaviour of a group of users because the information actions performed by individual users are widely variable, but their behaviour can be classified into a limited number of identifiable patterns (Sollund, 2007). In fact, Amazon -which uses recommendation algorithms to personalise the online store for each customer- uses collaborative filtering and cluster models (Linden *et al.*, 2003).

Nevertheless, in order to provide effective personalised assistance, information agents depend on the knowledge on individual users contained in user profiles. Therefore, an accurate representation of user interests is crucial to the performance of personal agents in satisfying user information needs (Godoy and Amandi, 2008).

In this line, this PhD dissertation uses a methodology for designing **ontologies** to describe the requirements of a virtual library that includes personalisation capabilities. The concept of an ontology has often been used as a synonym of taxonomy. In functional terms, they can be used as synonyms (Gruber, 1995), since computer

scientists use the term ontology whereas librarians utilize the term taxonomy (Adams, 2002). Anyway, there is a remarkable difference between these two knowledge organization tools as, although both of them allow the definition of concepts and relationships in a semantic way, the main users of ontologies should be computers, in spite of the fact that they must be human-readable as well. Hence, ontologies are intended to explicitly represent the objects, concepts and other entities that are assumed to exist in some context, together with the relationships that are held between them (Gruber, 1995; Qin, Paling, 2001). The definition for the term *ontology* in this PhD dissertation comes from the semantic web community (Heflin, 2004). Here, ontologies will be used to describe the requisites of a personalised system which uses all the relevant information for the process of searching and browsing a digital library to build a complete navigational profile for each user and its semantic description.

In order that the content and interface components can be personalized for a given user in a given context, we need to describe a set of key informational components, such as those defined by WC3 (2009):

- A functional description of personal needs and preferences; that is what the user requires in machine readable form.
- A functional description of the immediate delivery context.
- Metadata describing properties of a resource in a form that can match (or not match) the personal needs and preferences of the user for that context.

This PhD dissertation deals with these three informational components.

On the other hand, **metadata** is what facilitates users to discover and identify existing materials and distinguish between different resources. It also allows the personalising of information presented via the user information profile (Foster-Jones and Beazleigh, 2002). Metadata provides controlled and structured descriptions of resources through searchable access points such as title, author, date, location, description and subject, but can also provide either interpretative information on the potential education application of resources or include information on the relationship between the resources (Friesen, 2002). These two types of educational metadata are labelled as either authoritative or

non-authoritative (Recker and Wiley, 2001). While authoritative metadata is generally determined by librarians, non-authoritative metadata is more likely to be generated by the final users of learning objects[2]. In this context, it will be interesting to automatically produce non-authoritative metadata through ways such as tagging the navigational actions performed by students and teachers when they use a Learning Object Repository (LOR) in their search for some particular goal. Regarding personalization purposes, LORs can provide different services to users, from the typical searching/browsing functionalities to personal services such as keeping track of user interests based on the educational resources he or she has searched and downloaded. For instance, SMETE is a learning object repository for teaching and learning in science and technology at all levels, which includes a recommendation system based on past user interactions (Neven and Duval, 2002). Overall, this is an approach for "allowing content to speak to you" (Rosenfeld and Morville, 1998). Metadata is the primary key that links information architecture to the design of database schema as it allows the applying of structure and power of relational databases to the heterogeneous, unstructured environments of web sites.

Therefore, searching and browsing activities in a virtual learning environment can be a useful source of information on learner behaviour. The basic idea of this thesis is that the interaction with learning objects carried out by different user profiles can be stored in a structured way, and then shared for future users with similar necessities thus overcoming information overload. At the same time, the analysis of such interaction may also reveal interesting facts about the use of learning resources that can be added as new metadata to such resources thereby improving their overall quality. In addition, the analysis of content usage can reveal patterns and relationships which can suggest other ways to better structure, organise and provide access to that content.

Another relevant source of data for personalising information delivery systems is the **information behaviour of users**. Information behaviour focuses on people's information needs, their information seeking, managing, giving and use of information

[2] Learning objects, that is to say "any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning" (LOM, 2000) are the elements of collections on educational material mainly managed by libraries.

in the varied roles that make up their everyday lives. As Wilson (1981) points out, the role that an individual assumes in everyday life or in the working or social environment will be determined by how he or she behaves with regard to seeking information. The introduction of context by Dervin and Nilan (1986) shifted the focus on the research from the information system to the user, investigating how to understand situations, information gaps and use. According to these authors, the environment is an aspect that conditions the needs and the process of seeking and, hence, the meaning of information to people can only be comprehended when their **context** is understood.

However, studies conducted to date have been centred on analysing information behaviour in a specific context, without taking into account that people seek and use information in different spheres throughout the day -i.e. their professional role, as a student- (Allen 1996).

In the case of lifewide learning, where students pursue knowledge for either personal or professional reasons (Department of Education and Science, Ireland, 2000), it is obvious that there are several contexts where the information behaviour of one individual takes place, mainly in the academic, professional and daily life environments. Lifelong learning can therefore no longer be confined to a classroom or in a formal education setting, since it takes place throughout life and in a wide range of situations. Furthermore, learning should not discriminate between a place and a time to acquire knowledge or the place and the time to apply the acquired knowledge (Fisher, 2000). Taking into account the e-learning students of the Universitat Oberta de Catalunya, the majority of them are mature students (Sutherland, 1999), more than 85% are fully employed and at least half of them have family responsibilities. Having these characteristics in mind, some differences between educating adults and younger students rely on the accumulated knowledge and experience that can add to or delay the learning experience. Another distinct feature is that adult education is thought to be voluntary and therefore mature students are supposed to be more self motivated (Knowles *et al.,* 1998).

Nowadays, lifelong learning is seen as a key issue in the knowledge society, as it not only enhances social inclusion, active citizenship and personal development, but also

competitiveness and employability (EU Commission, 2006).

Accordingly, this PhD dissertation analyses the behaviour and reasons of students for their actions, not only for reaching their learning goals but also their professional and daily life aims. This will lead to a classification of users based upon their information behaviour which may provide useful clues for the design of information systems.

In our research, context can be observed as the environment where the adopted information behaviour is more clearly manifested by the displayed information-related competencies. Therefore, from the perspectives of the search process and the use of information, we will study how information competences are acquired in the different environments and which of these competencies acquired in a specific context are transferable to other contexts of action. Specifically, we will deal with the academic environment, which would encompass the activities related to the search and use of information in formal learning; the working context, which would include the activities related to the responsibilities derived from work; and the daily life context, which we understand as the information activities related to the tasks as being responsible for a family, as a member of the community, as a consumer, as well as the information activities related with leisure.

Besides, there is little research on the use of information and this existing research chooses a bibliographical point of view, using bibliometric analysis techniques, therefore excluding people who do not search or publish in academic journals (González Teruel, 2005). Our research aim is, therefore, to find out and understand the information behaviour and the information competences of a specific group, not only in the academic environment, but also in the professional and daily life environments and their interrelations.

In order to be able to study information behaviour, it is of interest to analyse the associated competences that should allow users to transform their knowledge, skills and attitudes into information actions in a specific sphere (Lloyd and Somerville, 2006). In a broad sense, **informationally competent** individuals are described as those who know when they need information and are able to find, evaluate, use and disseminate information effectively to resolve personal or professional problems or broader social

aspects (UNESCO, 2003). An informationally skilled user is that person who has adopted an appropriate information behaviour to identify, through whatever channel or medium, information well fitted to his/hers information needs, leading to the wise and ethical use of information in society (Webber *et al.*, 2005). People that are information literate are ready for lifelong learning because they can always find the information they need for the task or decision that they have to tackle (AASL, 1998). Information competences[3] are considered to be key and strategic for people to be able to fulfil themselves in life, to be active citizens, to feel part of society and to enter the employment market easily (European Commission, 2004) which are vital objectives in open learning scenarios (OLCOS, 2007).

Today, what is valued in the employment sphere is the ability to manage information needs effectively and there are studies that highlight the costs involved in having workers without information competences (Cheuk, 2002) or OECD studies that stress the rise in demand for informationally competent workers (OECD, 2000). However, studies since the 90s have found that students are leaving university "without the necessary transferable skills to cope in an information based society" (Ray and Day, 1998).

To sum up, there are no studies that focus on whether the consequences of informational incompetence can be palliated in the world of employment on the basis of a transfer of acquired competences, for example, from a formal university learning sphere[4]. Accordingly, this PhD dissertation addresses the question of whether the flow of transfer can go in this direction and help improve this type of situation.

The users which are the participants of the field work performed in this research are e-learning students from the **Universitat Oberta de Catalunya** (UOC, known as the Open University of Catalonia in English), a completely virtual campus which offers 19 official degrees, several graduate programmes and post-graduate studies, and a doctoral degree, with more than 40,000 students and more than 1,500 staff which include

[3] In the Commission's working framework, the working group chose to use the term *digital competence* when it wanted to refer to *information competence* as the skills it includes refer to information competences. These can be seen listed in the report of the OLCOS project (OLCOS 2007).
[4] The formal education system is hierarchical and chronological in structure, going from school to university (Coombs and Ahmed, 1974).

instructional designers, teachers, tutors and all other academic and technical staff. The UOC virtual campus is an integrated e-learning environment which allows users to communicate with other users using a mail system and includes an agenda, a news service, virtual classrooms, a digital library and other e-learning related tools. Although the use of classic text printed books is still massive, there is also a growing use of web based e-books and other online learning resources, so the introduction of new e-learning standards and application profiles such as IEES LOM and IMS-LD is becoming a necessity to maintain the constant development of the virtual campus. The increasing amount of multimedia content and standard based learning resources spur on a constant need for new functionalities and capabilities to be offered by the digital library.

As we have already stated, there is a significant number of studies into information behaviour in the academic sphere (Talja, 2005), and even in the case of distance students (Thósteinsdóttir, 2005), however, we have detected that the research to date has not tackled our research subject as we have proposed it. Our students are in a university that use ICT intensively and, at the same time, this is an open university, i.e. primarily addressed to adults and łifelong learning. Consequently, the profile of these students is rather different from those of face-to-face universities and their academic context is presumably more affected by their professional and private environments. The professional sphere is the least studied from the point of view of information behaviour and, according to Wilson, this is a pity as the contexts of use may be very different (Wilson, 2006). Therefore, besides the interest in our users in their role as students, we are also interested in their roles at work. Despite the fact that there are no consistent statistics on the application of e-learning, the OECD states that this online learning mode is a growing trend (OECD, 2005).

1.3. Work hypothesis and questions

In a virtual learning environment, students have to access and use a diverse but structured range of information (e-mail, news, calendar, learning resources, etc.) with a highly specific aim, namely the achievement of the learning goals of each training action in which they participate. There is a challenge for virtual libraries or any information delivery system embedded in a virtual learning environment to provide better support for information seeking and the use of information through personalised services and presentations, in order to increase their effectiveness and efficiency. A cost-effective way of obtaining data for personalisation purposes can be based on the previous navigation actions that users perform with learning contents. These contents are managed by each student according to his or her own organisational skills and habits, reproducing and applying information competences acquired in professional and private contexts, but also according to two intrinsic factors in the academic sphere: first, the technology (i. e. the information system) provided by the virtual learning environment and, second, the standardised training for incoming students that is part of the learning process. The virtual learning environment becomes, then, a common space where students develop their information behaviour applying the knowledge and skills acquired in any of the three spheres (professional, dailylife and academic) with specific aims. It is this transfer of competences that has a significant influence on the way that a student tackles the obvious need to treat and use all the information that is generated during the learning process and also the attitude that the individual places in each context.

In order to specify the objectives of the afore-mentioned research, we propose the following work hypothesis and the associated research questions.

Hypothesis 1: The navigation behaviour of the users of a virtual library generates information for offering personalised services.

a) How can the elements of a personalisation system of a virtual library be semantically represented by the data coming from users' navigational behaviour?

b) How can the information behaviour of e-learning students be reused for rating learning objects?

c) What metadata fields can be automatically filled from analysing user interactions with learning objects?

Hypothesis 2: E-learning students can adopt different personal information behaviour patterns in their academic, workplace and daily life contexts, while the information-related competencies acquired in any of the contexts are transferred to the other contexts.

d) What are the variables involved in the different patterns of information behaviour of e-learning students?

e) What are the personal information behaviour patterns of e-learning students in all their contexts of action?

f) Is it possible the transfer of information-related competencies from/to academic, daily life and workplace contexts?

1.4. Methodology

Research into information behaviour may, in broad terms, be organised between the system-oriented paradigm (from a traditional point of view) and the user-oriented paradigm (a newer alternative point of view) (Dervin and Nilan, 1986; González-Teruel, 2005). The former describes the user in terms of demographic variables that characterise the usage of a specific information system, while the latter aims to give importance to aspects such as the user's sphere or situation. From the point of view of the system-oriented paradigm, user profiles with similar characteristics needing the same type of information can be established. From the user perspective, the aim is to study how the user's sphere means that a set of variables, which are unique for each individual or group, shape information behaviour (González-Teruel, 2005).

The research methodology is also different according to the paradigm used. Generally speaking, under the system-oriented paradigm, questionnaires, analysis of quotes, references and records of interactions made on a specific information system are used. On the contrary, the user-oriented paradigm opts for more qualitative methodologies.

In the specific case of research methodologies and techniques in Information Studies, it is considered that the most suitable approach is a multiple methodology one (Glazier and Powell, 1992; González-Teruel, 2005). Specifically, in User Studies we decide not just to use qualitative methods but to complement it with a quantitative point of view (Wilson, 2006). "Qualitative data can be said to yield more detailed data – richer data", in other words, that together with the collected data provide descriptions on the activities performed throughout the phenomenon. Additionally, it seems a reasonable choice when multiple realities merge in the subject of research (Glazier and Powell, 1992), as it happens in this PhD dissertation where different contexts of action are under study.

In this sense, it is interesting to point out that in the first reported discussions on human behaviour in ancient times, the term "character" (what it can be understood nowadays as "behaviour") was defined as the imprint that craftsmen left on their work material. This definition was only focused on the ways of doing. It was the scholar Teofrast (pupil of Aristotle) who added the connotation of a way of being (Batalla, 2009) to the term.

Therefore, in this research work, we decided to approach the human way of doing with quantitative methodologies, while we addressed the way of being and the reasons for acting with qualitative techniques.

For these reasons, a multiple methodology approach was adopted. We used ontologies to describe scenarios of use of information delivery systems; log analysis for recording and studying the real navigation behaviour; in-depth interviews and content analysis for the comprehension of information behaviour. Hereunder, these methodologies are examined in greater detail.

a) Design of ontologies

The semantic web community defines an ontology as "a formal description of a possible scenario or context"; in other words, what exists can be represented by an ontology (Sheth *et al.*, 2005). Besides, from the point of view of knowledge representation, the most accepted definition is that "an ontology is an explicit specification of a conceptualization" (Gruber, 1993), or more accurately, "a partial specification of a shared conceptualization" (Domingue and Motta, 1999). Conventional knowledge organizational tools such as classifications and thesauri are similar, as they allow the definition of concepts and relationships in a semantic way.

Ontologies were used for describing all the richness of the possible usage scenarios of a digital library and the relationships which can be established among all the elements in such scenarios, pursuing the enhancement of information representation and an information retrieval system (Gruber, 1995; Qin, Paling, 2001).

The methodology for designing ontologies allowed us to approach the research question "a"[5] on how could the elements of a personalization system of a virtual library be semantically represented based upon the data extracted from the navigational behaviour of users.

The work presented in this PhD dissertation extends the two first steps for designing an ontology (Denny, 2002): namely, acquiring domain knowledge and organizing the

[5] a) How can the elements of a personalisation system of a virtual library be semantically represented by the data coming from users' navigational behaviour?

ontology. Regarding the first step, in order to assemble the appropriate information and defining all terms, several experts from the UOC (librarians, members of the department of educational technology and usability experts) were requested to get involved in a focus group session. The focus group was centred on discussing several possible micro scenarios of use of the UOC virtual library. Afterwards, to organize the ontology, the identification of the principal concepts and properties was established, and then the relationships among such concepts. Finally, the definition of actions performed between the concepts was stated.

b) Log analysis

Analysing the log data coming from the interactions of users with learning resources has provided insights about the real information behaviour of users, without the partiality of laboratory experiments which requires the active participation of users in an artificial environment. Users know in advance that their behaviour is going to be recorded and studied (Harley and Henke, 2007).

In this case, the experiment consisted in tracking the navigational activity performed by the students in the UOC Virtual Campus during one semester (February-June 2006), in order to analyse the level of interaction between students and the learning resources available in the campus. The study focused on two virtual classrooms; one from Computer Science (Statistics) and one from Information Science (Archives). Both subjects were mandatory and totalled 350 students.

The log files generated by the campus web servers were identified as a potential source of data regarding the interaction of the students with the learning objects. Those files were analysed using a custom-made software tool that reconstructed the navigational activity performed by the students in the virtual classrooms under study.

The raw data upon which this analysis was performed had been generated by the virtual campus web servers during an academic semester. By default, all HTTP requests received by a web server are recorded in a log file as a line of text that contains the IP address of the client, the timestamp of the request, the text of the request and, finally,

the text of response, typically a number codifying success or some error. For the period under study, those files contained the records of around five million HTTP requests representing all the navigational activity during that semester in the whole university, including all students of all subjects.

Briefly, the software tool reads the recorded HTTP requests from the log files and organized them within the campus sessions, based on their IP address and timestamp. Thereafter, it selected those sessions related to the classrooms of interest and, finally, produced a list with all the sessions of interest and another list with all the resources requested in those sessions.

More precisely, the process consisted of sequentially reading the request appearing in the log files and assigning them to sessions according to their IP and timestamp. If the IP matched that of a previously open session, the request was added to that session; if not, a new session was opened to handle that request. An open session is closed either when it was assigned a request for logout or when a predefined time-out had elapsed and is computed between the time of the last request in that session and that of the request currently being processed. Typically, all sessions started with a request for login into the virtual campus and most of them ended with a request for logout. Between these two requests, there could be a variable number of requests to resources from the classrooms, the library or other parts of the Virtual Campus. All the campus sessions occurred during that semester were reconstructed with this process and, obviously, not all of them were of interest for this study. To focus on the subjects under study, only those sessions that included some request on the resources of the selected virtual classrooms were selected. The output of this process was the list of the campus sessions with at least one request to the classrooms of interest, where each session had an IP address, two timestamps indicating the period during which this session was alive and a list with the resources requested during that session.

As the IP addresses were typically dynamic, the relationship between an IP address and a student within a session was ambiguous and consequently the student associated to each session was unidentified. This imposes a limitation on the functionalities that can be derived from this kind of log analysis. More precisely, this fact prevents tracking the

activity of an individual student during the course of his/her different sessions throughout the semester, as his/her activity is split in a large number of sessions with different IPs. It also prevents the matching of the recorded activity of the student with his or her personal attributes, curriculum, scores, etc.

From all the sessions, a list of accessed resources was generated, where each resource contained a list with all the requests on it, characterized by a timestamp and an IP address. This allowed us to analyse the number of requests to each resource and the evolution of those requests with respect to time throughout the academic semester.

The learning objects selected as examples for analysing the interaction of the users with them were those most needed, from the teaching plan point of view at the beginning of the semester. These objects were the "Erratum file" and the statements and solutions of the exercises. The focus was to detect the typical behaviour rather than the usage of individual users.

This methodology was chosen in order to address the research questions linked to the reutilisation of the information behaviour of e-learning students for rating learning objects (research question "b"[6]) and to which metadata fields could be automatically filled (research question "c"[7]).

c) Interviews

The data collection of the qualitative part of the research carried out in this PhD dissertation is based upon the performance of 24 semi-structured in-depth interviews conducted on students from a purely virtual university that does not include any face-to-face or blended activities (except final exams). The participants were mature e-learning students given that they were over 25 years old (Sutherland, 1999; Given, 2000) as well as being active professionals. In the course of 2007-08, there were a total of 40,264 students enrolled in undergraduate programmes, where the average age was between 26 and 35 years (53.8%), 55% of them had children, and more than 85% of the total

[6] b) How can the information behaviour of e-learning students be reused for rating learning objects?

[7] c) What metadata fields can be automatically filled from analysing user interactions with learning objects?

students had a full time job.

Keeping this typology of students in mind, a random selection was performed according to categories on a series of variables of prime interest for this research work. This selection was not proportionate, in the sense that the number of samples in each category did not necessarily correspond to their relative size in the population.

Three age groups were singled out: a group of 25-35-year olds which comprised the student average age; a group of 35-45-year olds and finally a group of more than 45-year olds, as this age is regarded as critical for the digital divide (Katz & Rice, 2002). The population was segmented in terms of gender (female/male) which can be considered proportional as there were 51% male and 49% female. Afterwards, we segmented each group in two subgroups that were called "Novice" and "Advanced" students. "Novice" was used to describe those students who had only been one or two semesters at the university, therefore they had some knowledge and skills of the virtual campus resources. On the other hand, we used "Advanced" for those who had been enrolled for 3 or more semesters and were supposed to be more familiar with the virtual environment and online learning.

Finally, we divided the groups once again in two subgroups "Experts" and "Non-experts" in terms of information-related competencies. "Experts" was used for students that were considered to be information competent and therefore fulfilled two conditions: the first one is that they had at least once, searched and retrieved an electronic article from any of the subscribed databases of the digital library and, secondly, they had, at least once uploaded content on the Internet, i.e. videos, photos, created a weblog, etc. The rest were considered "Non-experts".

The initial selection was designed through telephone interviews in order to hold face-to-face meetings. Prior to the face-to-face meeting, each student received a small questionnaire by email about when was the last time they searched and used information in the three different contexts. This step provided the interviewers with an incident case that would be explored during the interview in order to make participants recall past real information-seeking situations (Talja, 1999).

Twenty four face-to-face interviews were then performed, each of them lasting from 60

to 90 minutes, between September and October 2007, in Barcelona, Spain. These interviews were audio and video taped with the permission of the participants. First of all, two interviews were performed on a trial basis. The structure of the interviews was organised in two ways and tested regarding which was the easiest for the participant. In the first one, the questions for the three contexts were organised around the academic, workplace and daily life. In the second one, the questions related to the actions in which the information behaviour was manifested were classified according to access to information, treatment, integration, evaluation, creation and communication (Léveillé, 1997). The second option was chosen given that the first one appeared as being too repetitive with the same questions in the three contexts, and ultimately tried to avoid aspects such as the transfer of competencies between contexts, which was an important characteristic of our research.

Finally, the 24 interviews were transcribed in text form and used as the raw data for content analysis.

Both qualitative approaches, the interviews and the content analysis process, bear the research questions encompassed with the variables that determine the different patterns of information behaviour of e-learning students (research question "d"[8]), the information behaviour patterns of the e-learning students (research question "e"[9]) and the possibility of transfer information-related competencies from/to academic, daily life and workplace contexts (research question "f"[10]).

d) Content analysis

All the raw data were independently human-codified by two coders who did not take part in the recorded interview process, as ideally suggested by Krippendorff (1980). The codification process was performed with the qualitative research software NVIVO 7.0 (Richards and Richards, 1998; Barry, 2004).

[8] d) What are the variables involved in the different patterns of information behaviour of e-learning students?

[9] e) What are the personal information behaviour patterns of e-learning students in all their contexts of action?

[10] f) Is it possible the transfer of information-related competencies from/to academic, daily life and workplace contexts?

First of all, the attributes of each participant were coded following their personal attributes as summarized in Table 1.

Personal attributes	
Gender	Female / Male
Age range	25-35 / 35-45 / More than 45
Profession	Main professional occupation / Retired
Ongoing studies	Social Sciences (includes Law, Communication, Economics, Psychology) / Humanities (includes History, Languages) / Technology (includes Computer Science and Telecommunications)
Semesters	Novice / Advanced
Subjects	Number of subjects taken per semester
Information related competencies	Expert / Non-expert

Table 1. Codebook for the personal attributes of the participants.

The second codebook (Table 2) is composed of seven variables.

Study variables	**Actions**	**Contexts**
Information-related competencies	Access to information	Academic (AC)
	Management of information	Workplace (WP)
	Information uses	Daily life (DL)
Attitude	Access to information	AC
	Management of information	WP
	Information uses	DL
Cognitive approach	Access to information	AC
	Management of information	WP

	Information uses	DL
Internet use perceptions	Access to information	AC
	Management of information	WP
	Information uses	DL

Table 2. Codebook of variables studied and the contexts analysed.

The first variable of the second codebook is related with information literacy as after the phone interview, it was evident that the *information-related competencies'* attribute in Table 2 was not providing a real picture of the knowledge, skills and attitudes that elearning students had in the academic context. Therefore, a node was created that permitted to describe this attribute in greater detail (Table 2). Finally, this node was coded by means of the concept of Information-related competencies (IRC): the skills, knowledge, experience and behaviour that an individual needs to find, evaluate and use information effectively (Virkus, 2006).

The following two variables (*attitude* and *cognitive approach*) come from the literature review through a substantive codification (Charmaz, 2007). One variable was the attitude involved with the development of the information behaviour and information literacy. This concept is related to social components (in the sense of interpersonal/intergroup relational patterns) and with various personality traits such as instinct, need or arousal. The two possible values for attitude were Implicated and Uninvolved. Implicated were those who could be essentially stimulated, in the sense that they were more likely to learn or seek information because they were intrinsically motivated. On the other hand, uninvolved were extrinsically motivated people; all those who used to perform just for receiving some extrinsic reward or to avoid punishment (Dörney, 1994).

Furthermore, other personality aspects were also studied, such as the *cognitive*

approach (Disciplined/Chaotic). Those with disciplined approach tended to be meticulous, organized, methodical and focused on the process. Those with chaotic approach used to be anarchic, following their own rules and more focused on results. These aspects are considered potentially important in relation to information behaviour and information literacy (Palmer, 1991). In literature, the cognitive style is seen as important in relation with knowledge creation which influences information behaviour. But some studies related the individual differences to cognitive styles when database searches on the Internet and virtual environments came to play (Ford, 2000; Kim 2001). These studies focused on the seeking process and distinguished the explorative individuals in performing searches from the people who build their searches in a narrower stepwise fashion. In this dissertation, the cognitive approach takes into account a wider view on the variable as it can be applied not only to the seeking process but also to the use of information.

Variable 4, *Internet use perceptions*, come also from the literature review. It can be considered that there are three interpretative repertoires in relation to the perception of the use of the Internet (Savolainen, 2004). The enthusiastic repertoire emphasizes the strengths of the Internet, conceiving it as a great enabler or as a technology of freedom. Positive expressions such as fast, easy, interactive are favoured. The second repertoire, Realistic, gives the Internet no absolute priority but its value is seen to depend on its relative advantages in specific situations. Finally, the Critical repertoire is a standpoint with some reservations about the advantages brought about by the Internet.

Variables *Access to information, Management of information* and *Information uses* have their justification in the actions in which the information behaviour is manifested: access to information, treatment, integration, evaluation, creation and communication and in which context the action took place (Léveillé, 1997).

Finally, each variable was codified taking into account the context of use, such as academic, workplace and daily life. After the codification, in order to assess the reliability of the categorizations, the Krippendorff's alpha coefficient (1980) was computed through NVivo 7. Inter-coder comparisons and a comparative report were generated. Krippendorff recommends the α to be 0.80 or above, at significance level

0.05, while variables for which α is between 0.667 and 0.800 can be used only for drawing tentative conclusions. Hence, in this study, reliability is above the recommended level for the variables (*attitude* and *Internet use perceptions*) whilst slightly lower for the other two. Each of the four variables was considered as a separate 0-1 variable: 1 if the variable was present in the postings and 0 when the theme did not appear in the postings. The results of the α-agreement are shown in Table 4.

Global variables	**Reliability**
Information-related competencies	0.75
Attitude	0.87
Cognitive approach	0.71
Internet use perceptions	0.92

Table 3. Test of reliability.

The data coming from the codification was then used for setting probability statements about the relationship between concepts.

1.5. Summary of methodologies

In this section the methodology used in this dissertation has been presented. See Table 4 for a summary of the methodologies and in which article they appear.

Publication	Methodologies used
Paper I: "Towards a personalization in digital libraries through ontologies" in *Library Management Journal,* 2005, vol. 25, no. 4/5, p. 206-217.	Use of ontologies for describing scenarios of use of information delivery systems.
Paper II: "Enriching e-learning metadata through digital library usage analysis", in *The Electronic Library*, 2007, vol. 25, no. 2, p. 148-165.	Log analysis for recording and studying the real navigation behaviour.
Paper III: User Centered Design of a Learning Object Repository", in *Proceedings of Human Computer Interaction,* M. Kurosu (Ed.): Human Centered Design, HCII 2009, Lecture Notes in Computer Science, no. 5619, p. 679–688.	In-depth interviews and content analysis.
Paper IV: "La Gestión de la información personal en usuarios intensivos de las TIC", in *El profesional de la información,* 2009, vol. 18, no. 4, p. 365-373.	

Table 4. Summary of the methodologies applied in this PhD dissertation.

2. Main results

The summary of the results and the discussion of the four publications presented in this PhD dissertation are outlined following the hypothesis and the associated research questions. Although as an abbreviated version of the four papers, this section presents a bounded articulation of all the ideas, results and discussions provided in the articles. Furthermore, these articles are the product of the research conducted at a specific time but this research has evolved during the time and the process of writing the rest of the four articles. These reflections and new approaches are also included in this part.

In the first paper, the elements for building a personalised information delivery system integrated in an e-learning environment are defined. Information systems can offer personalised services adapting their content and display to different types of users. Two elements that constitute the user model determine the functionalities of the personalization system: firstly, user profiles, including navigational history and user preferences; and, secondly, the information collected from all navigational behaviours carried out by digital library users. The description of the relationships between these elements is provided using the design of ontologies for defining the requirements of a personalised information delivery system. Figure 2 shows a schematic diagram of the main ideas developed in this paper.

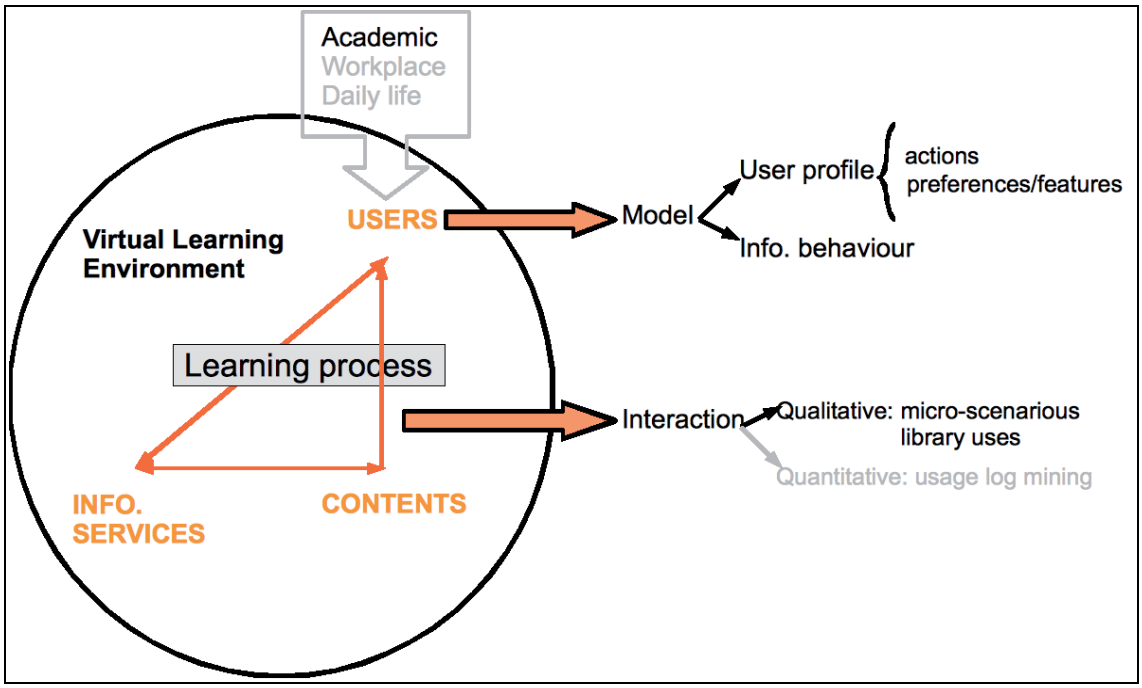


Figure 2. Schema of Paper I. (Elements in grey are not discussed in this paper and they are only shown for contextual purposes).

Hypothesis I is directly addressed in this paper.

Hypothesis 1: The navigational behaviour of the users of a virtual library can generate useful information for offering personalised services.

a) How can the elements of a personalisation system of a virtual library be semantically represented by the data coming from users' navigational behaviour? (Paper I and II)

First of all, the functionalities of the desired personalisation system were identified. Two basic elements were laid out: on the one hand, user profile, including navigational history and user preferences; and, on the other hand, the information collected from the navigational behaviour of the digital library users.

Following the basic procedure for building an ontology (Denny, 2002), the first two steps were covered, that is, the acquisition of the domain knowledge and the organization of the ontology.

The basic user profile' attributes are used to build the user model are shown in Table 5. For each attribute, it was determined whether it could be or should be explicitly given by the user; the relative importance for navigation; the recommendation system; and the user profiles which were more involved.

Attributes	Is it explicit?	Navigational support value level	Library recommendation level	Profiles concerned
Academic Register, Actual enrollment	No	High	High	Student
Academic Register, Previous enrollment	No	Low	High	Student
Navigational history	No	High	Medium	All
Navigational behavior and strategies	No	High	Medium	All
Socio-demographical profile, General background	No	Medium	Low	All
Socio-demographical profile, Academic background	No	Low	Medium	Student
Knowledge area	Yes	High	High	Teacher, Consultant, Researcher
Interests	Yes	High	High	All
Preferences	Yes	High	High	All

Table 5. Basic user profile attributes.

The formal model of navigational behaviour of digital library users is shown in Table 6. This model covered all the current information sources present in the digital library, the basic actions that could be taken by library users and the influence of those actions on the personalised information delivery system.

Navigation	Actions	Influence
Catalogue		
Searching	Search physic and analogical documents (books, VHS, DVD, etc.) into the OPAC	Low
Browsing	Navigate through author, subject, keyword, title, title course	Medium
Browsing	Look into the bibliographic register	Medium
Browsing	Look into the abstract and table of contents (for each document the DL scans the front cover, back cover, index)	Medium
Command	Loan command	High
Digital collection		
Browsing	Navigate among the thematic classification nodes	Medium
Searching	Search inside and electronic database	Medium
Command	Download, print, saving searches, sending searches by e-mail, create a Table of Contents (TOC) alert, add a journal to favorites, etc.	High
Subscription services		
Command	Subscription to the News Services (for each study the librarians create a weekly newsletter with a news selection about the subject of the study. Users received it as an e-mail)	Medium
Command	Subscription to paper journal TOC	High
Command	Subscription to SDI (Selective Diffusion of Information)	High
Library classroom		
Browsing	Look into the recommended bibliography	Medium
Browsing	Look into the FAQs	Medium
Browsing	Look into the didactic material	Medium
Browsing	Look into exercises of other semesters	Medium

Table 6. Information sources and basic user actions in the digital library.

The ontology is then built through crossing both tables (Table 5 and 6) by describing the relationships for recommendation purposes. Such a description can be created through the definition of small micro-scenarios (Young and Barnard, 1987) of typical uses of the library. For instance, suppose the following micro-scenario: John is a researcher working on his PhD on machine learning, as his profile says. He is working with Mary, his advisor, who got her PhD in the same field. Mary is also the advisor of two other students, Peter and Ann, who are also working on machine learning related subjects. When the students search for scientific papers, the results are sorted depending on the academic papers' relative importance, according to the number of times each paper has been downloaded by other researchers (giving more importance to the other students who work with Ann) but also on whether such a paper has been reviewed or not by any senior researcher (giving more importance to Ann's opinion). With the appropriate tools, such as Text2onto[11] and Protégé[12], and ontology definition languages, for instance OWL[13], such a statement can be translated into one or more rules that feed the

[11] http://ontoware.org/projects/text2onto/

[12] http://protege.stanford.edu/

[13] http://www.w3.org/TR/owl-features/

recommendation system with the actions taken by all the users in the digital library.

In spite of the fact that all members of a virtual community know in advance that the web based environment is logging all their actions (Carbó *et al.*, 2005), the recommendation system should be designed in a very friendly manner, including the possibility of disconnecting it like that of a non intrusive being. In addition, user actions are processed anonymously in the personalisation system and all the records are used only within the recommended system and not for any other reason. There is evidence that individuals are sometimes eager to be identified and desired to become part of a community (Lynch, 2003). Besides, as the information delivery system comes from an academic library and from a university, the trust of users in it is very high. Furthermore, the community built upon a university or a library is based on trust between the partners. Trust is an important issue in social navigation (Iivonen & Huotari, 2000). The feeling of trust is the first step in the promotion of the concept of social navigation (Davenport, 1999), subsequently, the personalised information delivery system can be applied, and thus, the mobilization of the expertise and the promotion of interactions among subjects can be pursued.

One of the strengths of libraries embedded in virtual environments is that "they combine content with community—community contribution, community involvement, and community annotation; they harness collections of people with common interests, not just content "(Lynch, 2003). The two elements of a learning environment, users and content, mediated through a personalised information delivery system, can enrich each another. Users through receiving recommendations and comments of users with similar interests our backgrounds and the collections by incorporating new content and deeper descriptions more attached to user interests.

The use of ontologies to represent user interests endeavours to close the semantic gap between the former approaches, which make use of low-level features extracted from documents (e.g. words), and the more abstract, conceptual views users may have of their interests. In spite of this advantage, committing oneself to a general ontology has a number of pitfalls. Firstly, when dealing with a large number of concepts, most general-purpose ontologies may become rather expensive for modelling a single user profile.

Secondly, ontologies often mirror the common-sense of a user community, but they fail to capture highly specific concepts that may be interesting for individual users, as well as the kind of documents users like to read in these concepts (Godoy and Amandi, 2008).

Summarizing, the results of the first research question show that the elements of a personalised system or a virtual library can be semantically represented by the data coming from users' navigational behaviour. The results provide, on the one hand, the basic user profile attributes to build the user model for the recommendation system and, on the other hand, the formal model of the navigational behaviour of digital library users. Finally, the user profile and the navigational behaviour model are crossed by means of the ontology which describes the relationships between them for recommendation purposes.

Paper II analyses the interactions between users and content. This article presents several experiments performed in a virtual learning environment, in order to propose an evaluation framework for analysing learning object usage, with the aim of extracting useful information for improving the quality of metadata used to describe such learning objects, but also for personalisation purposes, including user models and adaptive itineraries, as shown in Figure 3. This article reveals facts about the real usage of the learning resources which can be added as new metadata, overcoming information overload by means of proper recommendations, quality assessment and improving the learning process itself, when described through formal descriptions. A selection of IEEE LOM (Learning Object Metadata) fields is provided to enrich the description of learning objects, automatically in some cases, from the use of the information resources performed by users of an academic repository.

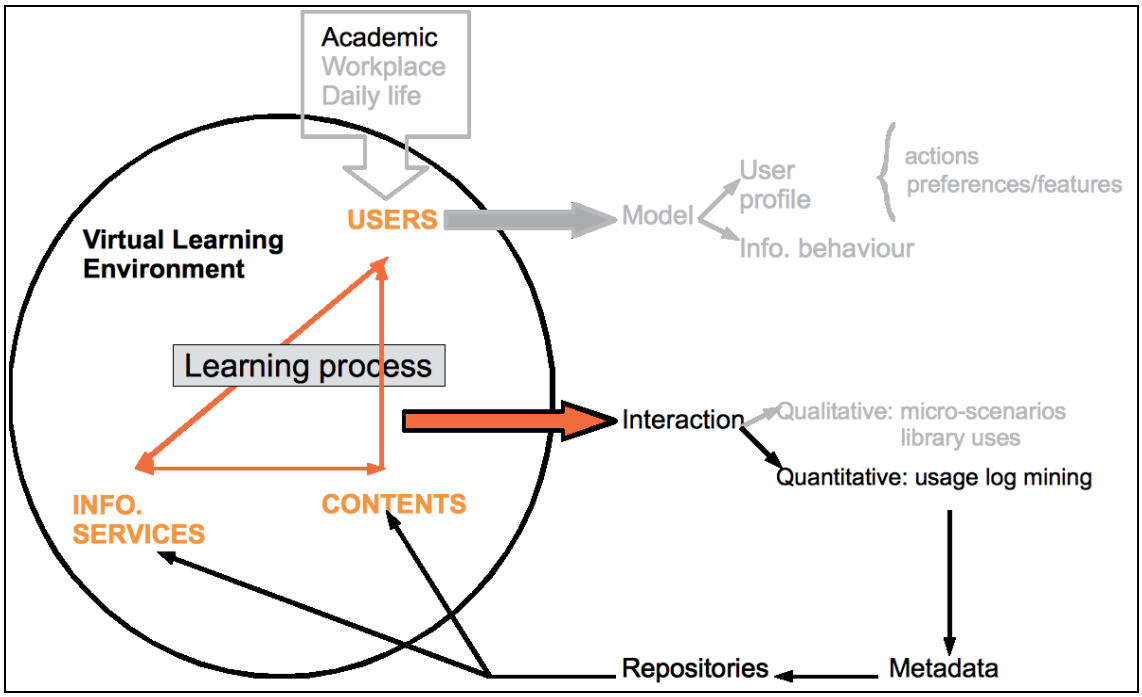


Figure 3. Schema of Paper II. (Elements in grey are not discussed in this paper and they are only shown for contextual purposes).

b) How can the information behaviour of e-learning students be reused for rating learning objects? (Paper II)

As presented in Section 1.4, an experiment was performed by means of tracking the navigational activity performed by e-learning students. During the second semester of the course 2005-06, the interactions between students and learning resources available in two virtual classrooms at the UOC virtual campus were analysed.

Log analysis allowed us to find, among all HTTP requests recorded by the Virtual Campus web server, those related with the classrooms of interest for this study. These were the requests for resources to the classroom plus the requests for other resources from the same IP in a reasonably short time. Then, all the web sessions containing one or more accesses to the classrooms of interest were reconstructed. Each web session was characterised by a timestamp corresponding to the login to the campus, another timestamp corresponding to the last request during that session, the IP address of the requester and a list of all resources requested during that session. The individual student corresponding to each session was not identifiable. Beyond privacy concerns, the IP

address was typically dynamic and, therefore, it would change among sessions of the same student. Eventually, through merging all the sessions of interest, we managed to reconstruct the story of actions regarding the resources related to the classroom, including any other resources in the campus that were accessed by the student in the same web session. This allowed us to characterise the interaction of students with each of the resources of the classroom in terms of the number of accesses to each resource and the distribution of those accesses along the time axis. Some observations and possible applications derived from these results are laid out below: resource usage, monitoring the learning process and identification of learning objects of interest outside the classroom.

Resource usage

As an example of practical application, a simple analysis of the accesses to each resource in the classroom revealed that some of them were never accessed. For those resources, there was no interaction with the learners. This may mean that users needed more metadata fields in order to know if a particular learning object is of their interest or that the title (the main access element) was insufficiently clear or drew their attention. For some learning objects, the lack of accesses could be explained by the fact that they are not cited in the learning plan, as some learning objects are neither included in the calendar nor in the prescription of the lecturers. Another possible reason may be that they were not considered relevant for the achievement of learning goals. Anyway, the information regarding the lack of usage of a given learning object can be useful when revising the teaching plan for further semesters. It also might be used as an alert system in order to warn the user that he/she has not used a resource that is important for the learning process.

Besides, the usage pattern of a given learning object can also be determined over time. As shown in Figure 4, there are learning objects with a specific life cycle which indicates that the resource is useful at a specific stage and irrelevant in other moments. It also indicates that the usage pattern is related to the calendar (holidays and week-ends), hence the usage analysis must be contextualised to the period of time linked to

the learning process.

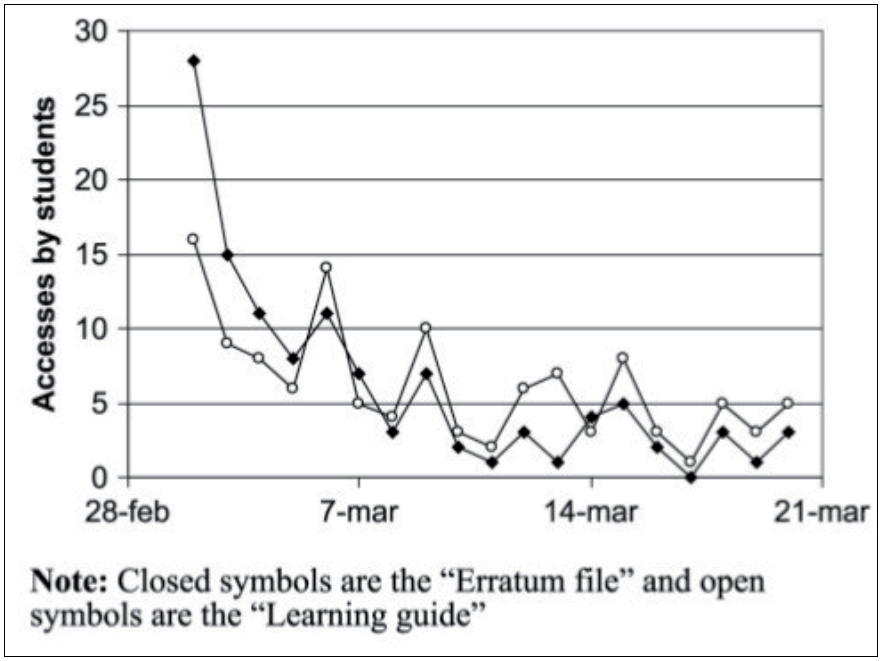


Figure 4. Dynamics of user's accesses to one learning object.

We have thus seen that navigational data show the relationships between user preferences and the learning processes in an implicit manner. A personalised information delivery system could be aware of those patterns over time and deliver recommendations in a way that supports this type of preferences.

Monitoring the learning process

Another application that can be supported with data derived from the log analysis is tracking the activity of students in order to verify if they follow the preplanned learning path. As an example, we compared the temporal patterns of accesses to statements of exercises with that of the accesses to the corresponding solutions. The data is represented in Figure 5. In this case, we observed that, students indeed followed the learning plan, as the accesses progressed as expected, and statements of the exercises were previously downloaded to their solutions, except for the last few days.

Therefore, we propose that using this type of analysis, behaviours that are not following

the learning plan could be detected before the final assessment is due, so there could be time for finding possible solutions if problems were faced.

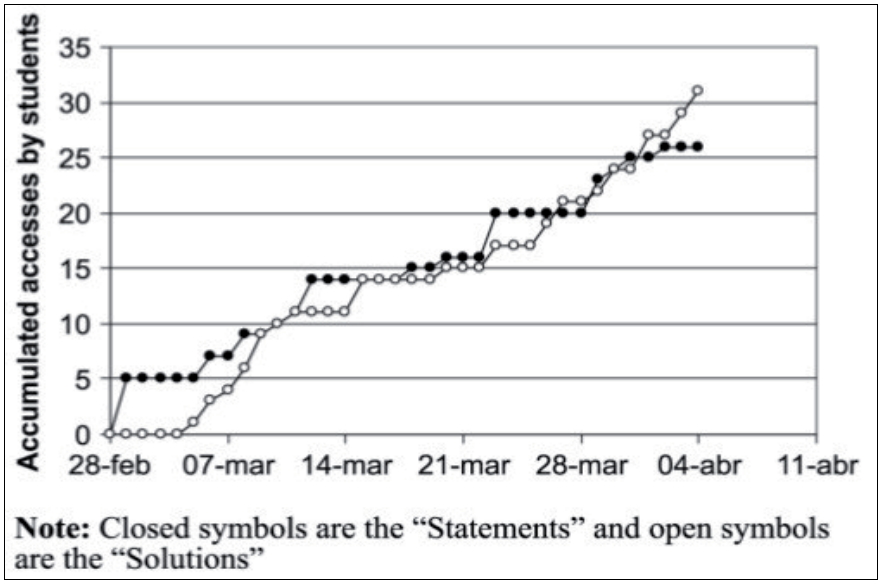


Figure 5. Dynamics of the interaction between learning objects and users in a virtual classroom.

Identification of learning objects of interest outside the classroom

Users did not only access resources of the classroom but, they also accessed the virtual library and navigated through it. Data from that navigation could also be recorded and analysed in order to identify what other resources may be of interest for students with a given assignment. This knowledge can be useful for further teaching planning if stored and structured in a proper manner (Sicilia and Garcia, 2005).

Overall, the procedure of log analysis provided teachers with information on the real information behaviour of learners without biases, for example, caused by laboratory studies. Participation of learners was not required but their behaviour was recorded passively, as other studies have also shown (Harley and Henke, 2007). However, the study of the information behaviour of use of students by means of log analysis is limited to the subset of actions described in Table 6 as logs logs do not provide data on the subsequent use that is made of the information in the documents.

Furthermore, users sometimes access information resources which are kept in servers

outside the library, that is to say, servers of information providers such as Elsevier, Springer or Thompson (and even open resources which are not subscribed by libraries such as Google or Wikipedia). It would be interesting if those providers could facilitate data derived from user navigation to the libraries, in order to reuse them for personalised services. From the UOC virtual library side, the only data that can be obtained was the access to such resources. As soon as users are inside an external server, all navigation actions are not recorded by the campus web server. Libraries and, more precisely, library consortiums when subscribing for electronic resources should take into account, at least, which usage data the provider should deliver periodically.

To summarize, this second research question is addressed by defining a basic system for analysing learning objects through the information behaviour of students in a virtual learning environment. This analysis can support the personalised information delivery system as well as the drawing up of learning plans.

c) What metadata fields can be automatically filled from analysing user interactions with learning objects? (Paper II)

Librarians describe the resources of catalogues, collections and repositories by providing authoritative metadata such as title, author, date, location, description and subject, which are used as access points. But another type of metadata called non-authoritative (Recker and Wiley, 2001) also exists which is likely to be indirectly generated by final users at a result of their navigational actions.

Following the Learning Object Metadata base schema (LOM, 2002), a selection of fields of non-authoritative metadata were selected to implement a process of metadata enrichment with information extracted from usage. More precisely, the LOM fields that were observed to be automatically generated through analysing usage are presented in Table 7.

Selection of LOM fields	Description	Enrichment of metadata
1.5. Keyword	Keywords or phrases describing the topic of the learning object.	Through the searches performed by librarians, teachers and students, it is possible to rank keywords according to the number of times they are used for retrieving a given learning object and, therefore, improving recommendations or detecting misplaced keywords, for example, as well as creating tag clouds for social tagging.
1.7. Structure and 1.8 Aggregation level	Underlying organizational structure of the learning object.	Usually, learning objects are described as independent chunks of information, as they are considered to have an atomic structure, supposed to be indivisible. Usage data may reveal relationships with other learning objects and this fact can be used to create collections, hierarchical, linear or networked structures.
3. Metadata	This field describes the metadata record itself.	It could be used for registering all the automatic changes that the system performs, in order to further analyse the metadata enrichment process itself.
5. Educational	This category describes the key educational or characteristics of the learning object.	Now, this is currently one of the most criticized aspects of learning objects and LOM, as it is clearly underused. It is related with the quality of learning experience, so it becomes critical for any intelligent tutoring system dealing with learning objects for building adaptive itineraries or providing personalised services.
5.1. Interactivity type and 5.3. Interactivity level	These fields can indicate active learning, expositive or mixed, and its degree.	These fields could be linked with the seeking information characteristic registered in the user profile. We could see, for example, that some resources provided in the "Statistics" course (additional exercises and examples) where not used by all users, as some of them preferred to base their study on the use of the hypertext material, which is more theoretical and textual, instead of practical. Other users prefer learning-by-doing instead and, therefore, under a personalized learning process they could be recommended to use learning objects with such interactivity type and level.
5.8. Difficulty	How hard it is to work with or	In an explicit way, users (the teacher but also the students) could suggest values for this field

	through this learning object for the typical target audience.	and the system could use it for selecting exercises to students according to their profile. This information can be also used by teachers to evaluate learning objects.
5.9. Typical learning time	Approximate or typical time it takes to work with or through this learning object for the typical intended target audience.	For pure online learning objects (i.e. exercises with embedded applets or simulations), the system can estimate the average time and use it to detect "outliers", that is, people that just walk through or people that spend too much time, taking the appropriate actions in each case.
7. Relation	This category defines the relationships between this learning object and other learning objects, if any.	This category can be used by the intelligent tutoring system to establish relationships between learning objects according to their usage, especially those detected from the adaptive paths followed by students.
9.Classification	This category describes where the learning object falls within a particular classification system.	As an extension of the concept of keyword, it is possible to analyse the usage of the terms in the taxonomy for discovering interesting relationships and unused terms.

Table 7. Selection of LOM fields involved in a process of metadata enrichment with information extracted from usage.

Some examples are provided, taking as case studies those learning objects analysed in the previous research question. Regarding the learning objects which were never used, their description could be improved through analysing the search terms used by students and modifying the LOM 1.5. field "Keyword", adding the LOM 5 "Educational" fields to such resources. In this way, these learning objects could be automatically recommended to users by using these enlarged descriptions. As for those learning objects whose use is observed to be closely associated, such as those corresponding to statements and solutions to exercises, this can be linked through the use of LOM 7 "Relation" field that defines the relationships between learning objects. Finally, for learning objects that show conspicuous time patterns of usage, such as the learning plan

and the erratum file, the use of LOM 1.7. "Structure" and LOM 1.8. "Aggregation" could allow a richer organisation of those resources and create collections based upon the relationships with other learning objects with a similar usage pattern.

Metadata are at the foundation of information systems in the sense that a metadata record is, after NISO, structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource (2004). Any navigation system, search tool, labels or any other element that supports information location and management has to be underlined by metadata (Lambe, 2007; Pérez-Montoro, 2010 *to appear*).

The generation of metadata within the information delivery system has to be performed focusing on the support of a certain level of human activity. As stated by Capurro and Hjørland (2003), we should not just regard our representations as objective, because this implies that we never fully specify the theoretical, social, and historical assumptions on which we act; all kinds of information systems have policies and more or less explicit goals, so that what we regard as information should also finally be a reflection of the social role of the information system. In this sense, the automatic generation of metadata fields through usage analysis may at least be a subjective way of aligning the learning goals of users with the aim of the learning object repository, namely promoting reuse.

Concluding, the personalisation of information delivery systems (i. e. learning object repository, a virtual library) can be improved through the proper exploitation of user preferences and navigational behaviour. Functionalities and requirements of the selected learning scenario are described through an ontology. Therefore, both user profile and navigation behaviour data can be automatically computed and finally stored in the metadata in order to be further used by the personalised system.

The aim in paper III is in learning object repositories but taking into account that the new European Higher Education Area (EHEA) promotes the design of learner-centred processes focused on the acquisition and development of competencies rather than on the consumption of contents, as shown in Figure 6. This is the reason why it is suggested in the paper that learning objects are more than systems for providing access,

reusability and preservation of contents. Instead, they should also be fully integrated into the learning process. In this sense, a qualitative methodology approach is presented as part of an information behaviour analysis as the background for integrating learning object repositories in virtual learning environments, following the requirements of a user-centred design (UCD).

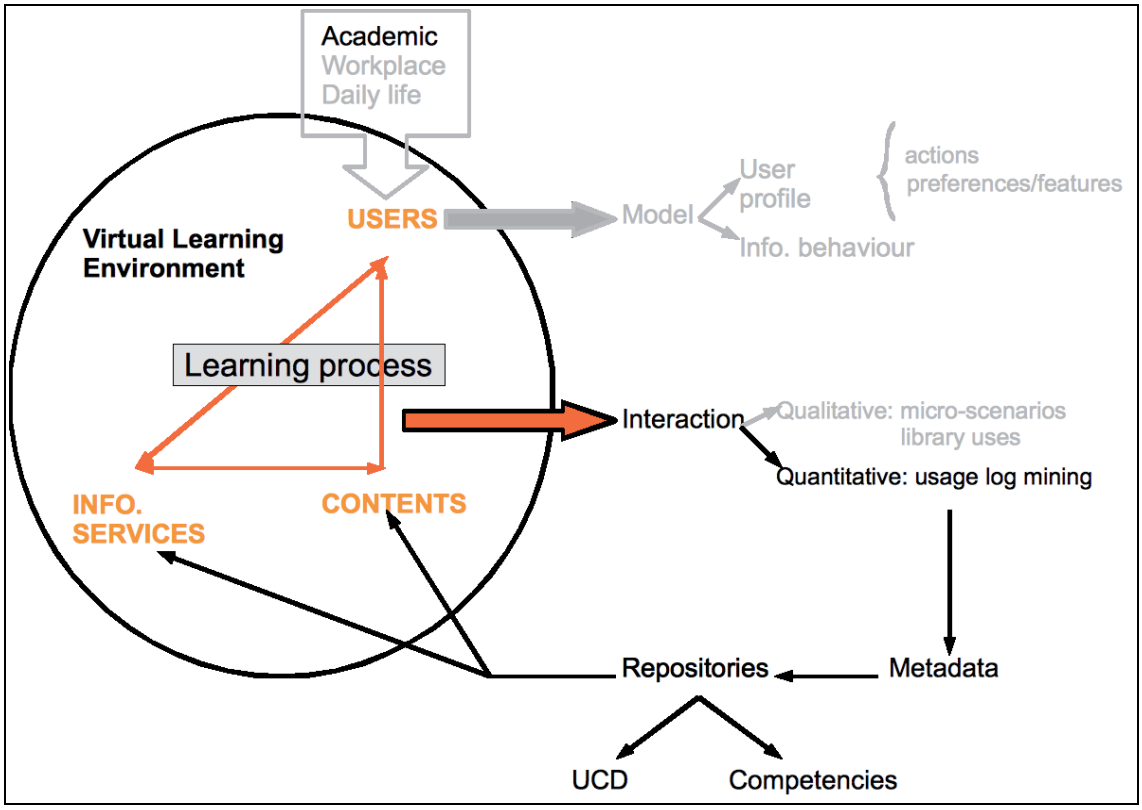


Figure 6. Schema of Paper III. (Elements in grey are not discussed in this paper and they are only shown for contextual purposes).

In paper IV, the qualitative analysis goes one step further and so it provides the identification of different user profiles as the first stage for designing a personalized information delivery system, as shown in Figure 7. Four user information behaviour patterns were identified, by combining two studied variables (cognitive approach and attitude): exhaustive, passive, reactive and proactive. These patterns show the differences in information behaviour for each context (daily life, workplace and

academic). Identification of user profiles is the essential first step in a user centred service design that addresses the specific intervention appropriate to each user type, taking into account the requirements of the tools and processes that users need to develop efficient information behaviour.

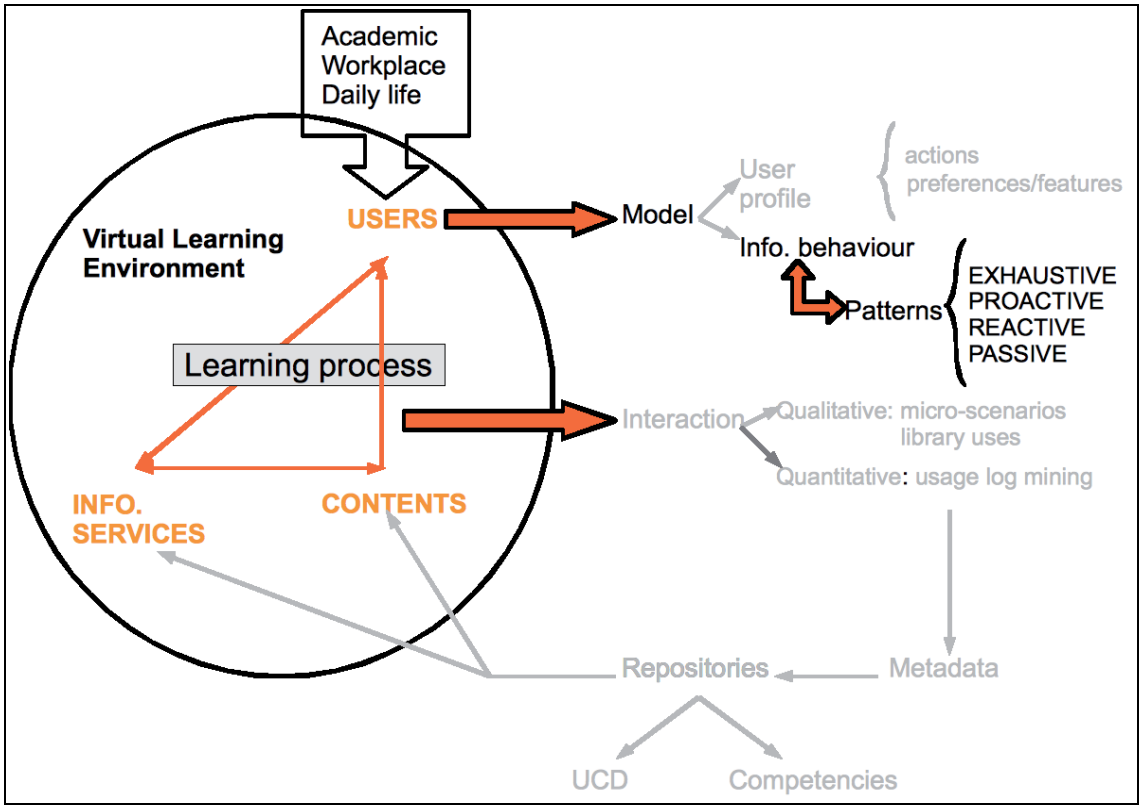


Figure 7. Schema of Paper IV. (Elements in grey are not discussed in this paper and they are only shown for contextual purposes).

Hypothesis 2: E-learning students can adopt different personal information behaviour patterns in their academic, workplace and daily life contexts, while the information-related competencies acquired in any of the contexts are transferred to the other contexts.

d) What are the variables involved in the different patterns of information behaviour of e-learning students? (Papers III and IV)

As explained in section 1.4, 24 in-depth interviews were performed on online mature students. The transcription of these interviews were codified using the following eight variables: Information-related competencies, Attitude, Cognitive approach and Internet use, Access to information, Management of information, Information uses and Contexts – i.e. academic, workplace and daily life-. As previously stated, these variables come from the literature. Additionally, some personal attributes were recorded, such as gender, age, profession, number of semesters in the virtual university, etc.

The results of this study show that four of the codified variables were involved in the different patterns of information behaviour of e-learning students. Those variables were Information-related competencies, Attitude, Cognitive approach, Internet use. Context was also a relevant factor.

More precisely, by crossing the variables Attitude and Cognitive style, the following four discourses could be established: Exhaustive, Passive, Reactive and Proactive. This classification comes from the clear relationship in the academic context between Attitude and Cognitive approach which produces four discourses with two opposing prevailing poles: Disciplined-Implicated and Chaotic-Uninvolved (as shown in Table 8).

Context	**Attitude/Cognitive approach**			
	Disciplined		**Chaotic**	
	Involved	**Uninvolved**	**Involved**	**Uninvolved**
Academic	9	3	3	9
Workplace	11	7	1	2
Daily Life	6	3	7	6

Table 8. Results of the correspondence between the C*ognitive approach* and A*ttitude* in relation to the context.

These two relationships (Disciplined-Involved and Chaotic-Uninvolved) can be explained through the fact that motivated people invest more effort in the process, resulting in a more disciplined strategy than discouraged ones, which aim directly at results, without any real interest in the learning process, thus leading to a more chaotic way.

Several authors have discussed about which variables are involved in the process of information seeking. The model of information behaviour of Wilson (1981) shows how psychological, demographical, role-related, interpersonal, environmental and source-related characteristics influence the information-seeking process. But as stated in the literature review, the previous studies are more focused on the seeking process in itself without taking into consideration the whole information behaviour. Instead, the focus on all contexts of action in this study allowed us to identify the variables that may affect the development of a particular behaviour in one context. For instance, in the academic environment a very structured and firm shape of the learning plan may lead to uninvolved attitudes, while the more open and flexible settings such as the daily life context, information behaviour presents more involved attitudes, which lead to a better level of information literacy.

Meanwhile, aspects such as gender, number of semesters involved in a virtual campus, type of studies currently undergoing, age and family responsibilities were surprisingly not significantly different with respect to those related to information behaviour. Usually, other quantitative studies based on library surveys provide some differences between age or gender (Rowland and Nicholasa, 2008) and even qualitative studies using in-depth interviews and focus groups (Waldman, 2003) suggested similar differences. The absence of differences in our results when compared to the extensive literature on this subject can be explained by the the special features of the participants of the research of this dissertation, as all of them were e-learning students and, therefore, all of them shared an enthusiastic approach regarding the use of ICT.

e) What are the personal information behaviour patterns of e-learning students in all their contexts of action? (Papers III and IV)

Despite the large number of variables influencing user information actions, ultimately, only a few variables were required to express individual behaviour to a relatively small number of common information behavioural patterns. As outlined in the previous research question, the variables Attitude and Cognitive approach allowed us to classify the interviewed students into four patterns, as shown in Figure 8, according to the

following codification:

- Exhaustive: disciplined attitude and implicated cognitive approach.
- Passive: chaotic attitude and uninvolved cognitive approach.
- Reactive: disciplined attitude and uninvolved cognitive approach.
- Proactive: chaotic attitude and implicated cognitive approach.

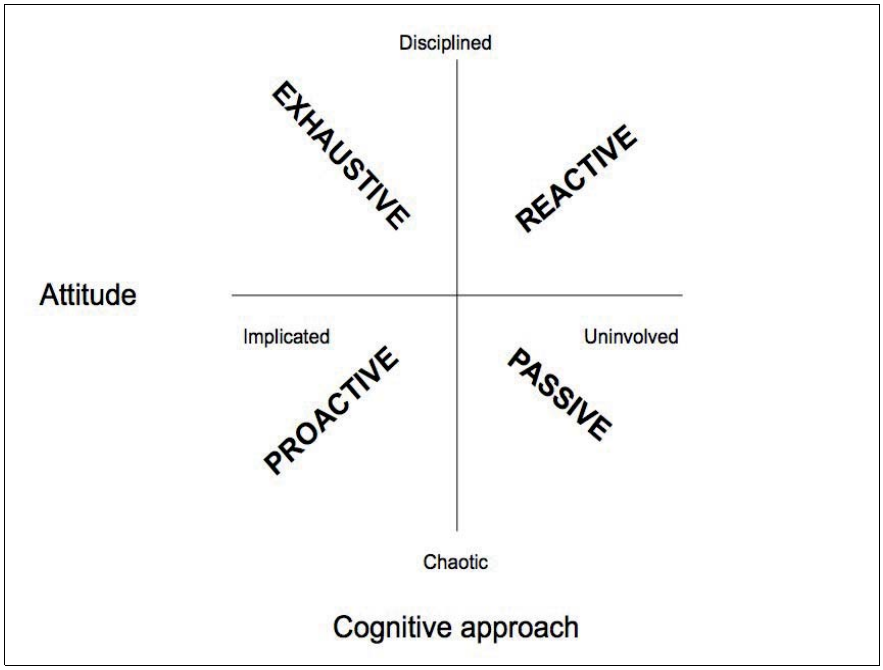


Figure 8. The four information behavioural patterns that appear when crossing the variables Attitude and Cognitive style.

Exhaustive: Disciplined / Implicated

Users classified under this pattern appear to be active when searching, managing and using information, finding these activities attractive as well as discovering new techniques and tools to improve these processes. When they learn or improve a certain skill in one particular context, they are very talented in transferring it to the other contexts. On the whole, their level of information literacy is at the same level in all

their contexts of action.

For them, it is essential to receive a roadmap; they appreciate guidelines and knowing the process in advance to get the expected results.

Here we provide a few statements, expressed by the participants themselves, to show this pattern:

"*As I have a lot of work to do at the UOC, I follow what the teacher says regarding the didactical plan and the recommended bibliography. I have discovered the trick that if I follow the marked steps I am successful (...), I always go to the face-to-face meetings (...), I have created and I keep a bibliographic database for study purposes*".

There is clear evidence that most of the cases are Disciplined-Implicated in the professional context as the disciplined attitude is the most prevalent in the workplace context.

Passive: Chaotic / Uninvolved

This pattern corresponds to users that are inconstant with procedures or that need to adapt their internal rules to the regulations of each context. They have an inclination to associate certain tools and processes to determine contexts. They need an environment that allows them to choose from a wide range of processes to reach similar results and to integrate tools to support the seeking and use of information.

A typical example could be "*I am very disorganised as I always forget to save the links I am interested in (...), I never participate in the classroom debates (...), I receive the didactical printed material at the beginning of the semester but I do not even open the package. (...), I prefer the option of PDF materials because then I can use the search engine of Adobe and find the answers of the questions of the teacher straightaway*".

Reactive: Disciplined / Uninvolved

The students classified under this pattern need to feel secure with the tools and that they have the knowledge and skills to start applying improvements in personal information

management. They happen to transfer processes between contexts, but they use the predetermined tools offered by each context.

"I use the resources of the virtual classroom only if the lecturer told me that they are useful for the evaluation".

"(...)I learned Excel and Access at work but then I started using them at home, but less intensively".

"*In all situations, the didactical materials of the virtual classroom are useful and there is no need to go into greater depth in order to get other educational resources*".

"*I give total priority to my work and the courses are of secondary importance*".

It is remarkable that some of those reactive in the academic context recognized that they were involved in their previous first face-to-face university experience, whilst in this second virtual university experience, they have a different Attitude: "*I used to make summaries and schemes when I studied at university, but not now (...). Now I have all the* didactic *material at the beginning of the semester and I do not need to search for more information*".

Proactive: Chaotic / Implicated

Finally, these students apply different processes and tools to different contexts as a personal feature. They use the newish tools and search if they are more open, flexible and agile.

"For my personal use, I use del.icio.us, Google Reader, bookmarks… but at work I follow the sets of protocol strictly which are clearer and more important to follow".

"I do not maintain a directory because the Internet is already the directory".

"I have so many things in the bookmarks that I couldn't find anything which is why I decided to create a database".

Overall, the classification of e-learning students into these four patterns indicates the relevance of Attitude and Cognitive style to describe the difference between information behaviour as well as the relevance of the context. In literature, studies do not focus in

the other contexts of action for a given individual. For instance, usability of e-commerce virtual environments usually focuses on the user as a consumer, without taking into consideration that the same person develops other actions in other contexts. These findings can bring about some suggestions for the design of virtual environments or, in the case of an academic context, for designing learning plans. Knowing how an individual informationally behaves in one context can suggest ways of displaying or providing content in the other contexts of action. This fact can improve user satisfaction with respect to the information delivery system, but can also contribute to the acquisition or enhancement of information-related competencies.

In short, four user patterns were established taking into account the information behaviour of e-learning students in their academic, workplace and daily life contexts. Aspects such those related to personal information management and related variables were analysed. The identification of user patterns was the essential first step in a user centred service design that addresses the specific intervention/contribution suitable for each user type.

f) Is it possible the transfer of information-related competencies from/to academic, daily life and workplace contexts? (Papers III and IV)

Apparently, common sense tells us that if a person is information competent in one context, he/she will develop an information competent behaviour in the other contexts in which he/she is involved. Our results indicate that this is generally true but only up to a basic level of expertise; for example, when searching and retrieving information. On the contrary, in a more advanced level of information competencies such as knowing the available information resources for each situation (and which of them are the most reliable) and how to manage them, we observed that information behaviour depended on the context. In this sense, some information-related competencies appeared in only one context and were not manifested in the other contexts. For instance, experts in workplace and daily life contexts became non-experts in the academic environment.

Applying quantitative techniques (namely Fisher's Exact Test) to the codified data derived from the interviews, we observed that the daily life and workplace contexts had

a strong correlation in terms of information-related competencies ($p<0.001$), but there were no clear correlations between them and the academic context. This suggests a relationship between daily life and workplace in the level of information-related competencies, while the academic context seems to be kept aside of this relationship. The lack of standards for competencies in both daily life and workplace contexts makes the determination of similar levels much more difficult.

In addition, the results suggest that the Cognitive approach is a variable that is not relevant for behaving as expert or non-expert in information-related competencies in all the three contexts. However, our data suggest a regular pattern involving cognitive approach and attitude. If participants were implicated (Attitude), they also showed a disciplined Cognitive approach, while if they were uninvolved (Attitude), their informational behaviour could be described as chaotic. Besides, in the academic context, the low level of information literacy was related to an uninvolved attitude with chaotic features, while in the workplace being implicated and disciplined meant being an IRC expert.

This difference in the academic context with respect to the implicated attitude and information expertise can be explained by the specific characteristics of the learning environment. The academic environment not only provides students with content but with a complete learning plan. Therefore, in this context, the processes are very clear and there is less opportunity for being proactive. On the other hand, in the daily life context, for example, there are multiple potential information needs and there is not a pre-established plan for the actions related with information.

This particular feature of the learning environment can be an opportunity for improvement in educational institutions. If it is clear then these information-related competencies need to be acquired and developed by learners, then the learning plan must include activities that foster a proactive attitude and, therefore, not only should provide learners with the required learning content but promote also the skills for finding, evaluating, using and communicating information. In this sense, the use of the learning object repositories becomes an important issue in the design of virtual learning environments. Hence, the learning plan should provide the knowledge as well as the

motivation for exercising the information competencies essential for living in the information society.

In this sense, our results bear out that while attitude was linked to the individual, as previously suggested, the cognitive approach depended on the context of action. Students exhibited a different cognitive approach when applying information behaviour in each context. For instance, it was the disciplined approach that was the most predominant in the workplace context, as it is an environment where professionals have always to behave in accordance with pre-established rules and processes. The goals in professional life require the following of a precise methodology in order to get the expected results. In this sense, disciplined experts were the most abundant while the chaotic experts were fewer. However in non-expert cases, the distribution was more balanced.

The findings also highlighted a transversal feature between the three analysed contexts. The Cognitive approach was not relevant in order to behave as expert or non-expert in information-related competencies. In the academic context among non-experts, both chaotic and disciplined were equally abundant. A possible explanation for this might be the low or null level of information competencies and, therefore, it suggests that information behaviour is neither determined by the methodology associated to the search nor by the management and use of information processes.

Similarly, it was found that the daily life context offers a wide variety of potential needs, different formats, a great variety of information resources and management tools. In this context, there is not a clear relationship between the cognitive approach and information-related competencies. A possible reason for this could be that in which the applying of a type of methodology for searching and managing information in this context may not be sufficiently rewarding if similar goals can be obtained from trial-and-error approaches or serendipity. A chaotic cognitive approach in the daily life sphere might be more sustainable than investing time in following a methodical plan. Planning does not directly bring about the satisfaction of information needs, as we cannot see a strong link between IRC expertise and the cognitive approach.

Regarding the aspect of transversality of key variables across the different contexts, in

the case of attitude, half of the participants had an implicated attitude in the three contexts. Analysing the remaining subjects of the sample, those who had different attitudes (i.e. depending on the context) were found to be implicated in the professional and daily life contexts, while they were uninvolved in the academic context. Nevertheless, for the whole sample there was a strong relationship between the attitudes in the three contexts ($p<0.001$), so being Proactive or Reactive in one context determined the attitude in the other two.

As previously stated, results indicated clear differences among the contexts in cognitive approach (i.e. either disciplined or chaotic). While a disciplined style prevailed in workplace, a chaotic style predominated in daily life and both styles were equally abundant in the academic context. We observed a relationship between the academic and daily life contexts ($p=0.017$), while the workplace context seemed to be independent of the other two ($p=0.248$ and $p=0.094$ respectively).

Finally, for the variable of Internet use perception, almost all the cases were coincident in being enthusiastic. In this case, the Fisher's exact test revealed that the three contexts were correlated when compared in pairs (Academic vs Workplace, $p<0.001$; Academic vs Daily Life, $p=0.003$; Workplace vs Daily Life, $p<0.001$). Hence, the studied population had as a common feature their enthusiasm for ICT, which was largely seen as a means for empowerment or liberation. This is consistent with the fact that all of them had chosen to study in a virtual university that uses Information and Communication Technologies intensively.

Regarding the relationship between the feelings about Internet use and the other variables; being an enthusiast was strongly related to being information-related competent in professional and daily life contexts (see Table 6). But contrary to expectations, this relationship was not observed in the academic environment. The reason for this may have something to do with the virtual learning environment where the tools for searching, managing and using information are predetermined by the virtual campus and the library and that educational content is canned into the virtual classroom and the didactic plan. It is unlikely that such a prearranged set of information needs, tools and resources is found at the workplace and it is certainly absent in daily

life, where people demonstrated that they used a great variety of information sources and tools, especially social software. Overall, our results about Internet use lead us to state that the Enthusiasts of ICT at the workplace and daily life are experts in information literacy in all contexts (See Table 9).

Context	IRC/Internet use					
	Expert			Non-expert		
	Enthusiastic	Realistic	Critical	Enthusiastic	Realistic	Critical
Academic	4	1	0	13	3	3
Workplace	12	4	0	5	0	3
Daily Life	12	3	0	6	1	2

Table 9. Results of the correspondence between information-related competencies and feelings regarding Internet use.

Those that were enthusiastic about *Internet use* were also enthusiastic in their expressions. A retired person said "*I am hooked on the Internet".* Another addict said "*I have become dependent on the theatre forums*". Another enthusiastic participant who had been forbidden to use the Internet at work said that at home her computer was "*exuding smoke*" as she was addicted to YouTube. Another way to express this enthusiasm but in this case in relation to e-learning was: "*Studying online gives me more freedom*".

3. Conclusions

In this dissertation, the information behaviour of e-learning students has been analysed in order to better understand the learning process in a virtual learning environment and the implications in the design of such environments and learning plans. As a central part of the learning process, the information-related competencies have been studied. As shown, these competencies as well as the other variables studied provide us with useful information for designing personalised information delivery systems. Due to the nature of mature students, not only the academic context but the workplace and daily life contexts have been considered to create a comprehensive vision of information behaviour. From this analysis, the following conclusions can be stated:

a) Using all the information relevant to the process of searching and browsing, the elements that determine the requisites of a personalised virtual library embedded in an e-learning environment are the following *(Paper I)*:

- User profile: including personal information, user preferences and navigational history
- Navigational profiles: data collected from the navigational behaviour of the digital library users

b) Regarding the resources involved in the learning process, they can be rated in an automated way based on registering the navigational actions of students. In this way, their information behaviour within each session in the virtual campus can be analysed and the access to each particular learning object can be studied *(Paper II)*:

- Even without the special recording utilities in the virtual campus, the interaction of students with learning objects can be tracked by analysing server log files where all HTTP requests are recorded by default.
- More advanced potentialities are limited by the anonymity of the students. In

this sense, a more in depth analysis would require recording an unambiguous identification of the student in each session, not just his or her IP address, and dealing with privacy concerns.

c) From the evaluation of the interactions between users and contents, a selection of Learning Object Metadata (LOM) fields can be automatically generated and be further used as non-authoritative metadata for rating those learning objects. *(Paper II)*.

d) As the first step towards the design of a personalised information delivery system, a User Centred Design (UCD) approach can be applied for planning and developing a learning object repository *(Paper III)*:

- Interviews and log analyses are complementary methodologies for gathering user requirements to obtain a service adapted to user profile and context of use.

e) The personal information behaviour of e-learning students can be classified into four patterns taking into account their attitudes and cognitive approaches (Paper IV):

- This user categorisation can be stored in the user profile to feed the personalised information delivery system.
- Other analysed variables such as gender, number of semesters of enrolment with a virtual university and family responsibilities do not determine differences in the personal information behaviour of e-learning students.

f) There is some transfer of information-related competencies between workplace, daily life and academic contexts, but this is only at a basic level of information literacy expertise *(Paper IV)*:

- The cognitive approach (disciplined/chaotic) is not relevant for behaving as expert or non-expert in information-related competencies in all of the three

contexts.

- The relationship between information-related competencies' expertise, attitude and cognitive approach is different for each context:
    - In the workplace, being expert in information literacy is associated with having an implicated attitude and a disciplined cognitive approach.
    - In daily life, the information-related competencies do not show any clear relationship with the cognitive approach.
    - In the academic context, the observed low level of information literacy seems to be related with a prevalent uninvolved attitude with chaotic features.
- There is a regular pattern involving the cognitive approach and attitude. If people are implicated, they will also develop a disciplined cognitive approach, while if they are uninvolved, they behave informationally chaotic.

As shown, the traditional way of providing content through virtual learning environments does not promote the development and acquisition of information-related competencies. It has also been stated that involved and disciplined features in learning environments are a key factor for acquiring such competencies. This new learning paradigm is aligned with the European Higher Education Area requirements.

Current and future research in this topic should include the development of learning object repositories as a key tool for providing not only content but true learning experiences for promoting the acquisition of discipline specific competencies as well as transversal competencies such as the information-related ones. Studies are needed about the integration of this new concept of repositories into the learning process through predesigned activities that encourage its use.

New trends in the design of future learning environments should take into account all the contexts where information behaviour is developed. Virtual Learning Environments should evolve towards a collection of services that users can aggregate according to

their preferences and special features. With this new concept of service aggregation, users will be able to overcome the boundaries between contexts and, therefore, having a unique Personal Learning Environment where workplace and daily life contents and skills can be merged.

Further studies need to be performed on information literacy in workplace and daily life contexts, similar to those in the academic area. In higher education studies, standards unify ways of analysing the knowledge, skills and attitudes of users, therefore helping to produce a more in-depth research in this subject.

4. References

AMERICAN ASSOCIATION SCHOOL OF LIBRARIANS (AASL). *Information literacy standards for students learning* [online], 1998. Available from < http://www.ala.org/aasl/ip_nine.html> (cited 15th of September 2009).

ABELL, A.; OXBROW, N. *Competing with knowledge: the information professionals in the knowledge management age.* London: Library Association Publishing, 2002.

ADAMS, K. "The semantic web: differentiating between taxonomies and ontologies", *Online*, 2002, vol. 26, no. 4, p.20-23.

ALLEN, B.L. *Information tasks: toward a user-centered approach to information systems*, New York: Academic Press, 1996.

BARLETT, J. C.; TOMS, E. G. "How is information used? Applying task analysis to understanding information use" [online], *Data, Information, and Knowledge in a Networked World Conference. Proceedings of the Canadian Association for Information Science*, held at The University of Western Ontario, Ontario, 2 - 4 June 2005. Ontario, 2005. Available from <http://www.cais-acsi.ca/proceedings/2005/bartlett_2005.pdf> (cited 15th of September 2009).

BARRY, R. "NVivo 2.0 and ATLAS.ti 5.0: A Comparative Review of Two Popular Qualitative Data-Analysis Programs", *Field Methods*, 2004, vol. 16, no. 4 (cited 15th of September 2009).

BATALLA, J. (ed). *Teofrast, caràcters*. Santa Coloma de Queralt: Edèndum, 2009.

BAWDEN, D. "Users, user studies and human information behaviour", *Journal of Documentation*, 2006, vol. 62, no. 6, p.671-679.

BOON, S.; JOHNSTON, B; and WEBBER, S. "A phenomenographic study of English faculty's conceptions of information literacy", *Journal of documentation*, 2007, vol. 63, no. 2, p. 204-228.

BRUCE, C. *The seven faces of information literacy*. Adelaide: Aslib Press, 1997.

BRUCE, C. "Information literacy as a catalyst for educational change: a background paper" [online], U*.S. National Commission on Libraries and Information Science, and the National Forum on Information Literacy* [online]*, for use at the Information Literacy Meeting of Experts*, held at Prague, The Czech Republic, 2002. Available fom <http://www.nclis.gov/libinter/infolitconf&meet/papers/bruce-fullpaper.pdf> (cited 15th of September 2009).

BUNDY, A. (ed). *Australian and New Zeland information literacy framework: principles, standards and practice* [online], 2nd edition, ANZIIL University of South

Australia Library, 2004. <http://www.anziil.org/index> (cited 15th of September 2009).

CAPURRO,R.; HJØRLAND, B. "The Concept of Information", *Annual Review of Information Science and Technology* , 2003, vol. 37, p. 343-411.

CARBÓ, J. M.; MOR, E.; MINGUILLÓN, J. "User navigational behavior in e-learning virtual environments", *Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)*, held at Compiègne University, France, 22-25 September 2005, p.243-249.

CASTELLS, M. *La era de la información. Economía, sociedad y cultura. La sociedad red, v*ol. 1, México DF: Siglo XXI Editores, 1999.

CASE, D. O. *Looking for information: a survey of research on information seeking, needs and behaviour*, Academic Press, San Diego CA, 2002.

COMMISSION OF THE EUROPEAN COMMUNITIES. *Adult learning: it is never too late to learn* [online]. Brussels, 2006. Available from < *http://eur-lex.europa.eu/LexUriServ/site/en/com/2006/com2006_0614en01.pdf>* (cited 15th of September 2009).

CORRALL, S. "Key skills for students in higher education", *SCONUL Newsletter*, Winter 1998, no.15, p. 25-29.

CHARMZAZ, K. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis.* Thousand Oaks, CA: Sage Publications, 2007.

CUNLIFFE, D. "Developing Usable Web Sites - A Review and Model", Internet Research, 2000, vol. 10, p. 295-308.

COLKIN, E. "Personalization tools dig deeper", *Information Week*, 2001, no. 852, p. 49-51.

COOMBS, P. AHMED, M. *Attacking Rural Poverty*, Baltimore: The John Hopkins University Press, 1974.

CHEUK, B. *Information Literacy in the workplace context: Issues, best practices and challenges* [online], White Paper prepared for UNESCO, the U.S. National Commission on Libraries and Information Science, and the National Forum on Information Literacy, for use at the Information Literacy Meeting of Experts, Prague, The Czech Republic, 2002. Available from <http://www.nclis.gov/libinter/infolitconf&meet/papers/cheuk-fullpaper.pdf> (cited 15th of September 2009).

CHOO, C. W. *Information Management for the Intelligent Organitzation: The Art of Scanning the Environment*. Medford, NJ: Information Today, 2002.

DAVENPORT, E. "Matching partners: Tools for social browsing that take people as texts", *Canadian Journal of Information and Library Science* , 1999, vol. 24, p. 1-16.

DENNY, M. “Ontology building: a survey of editing tools”, *XML.com*, 2002, Available from <http://www.xml.com/pub/a/2002/11/06/ontologies.html> (cited 15th of January 2005).

DEPARTMENT OF EDUCATION AND SCIENCE, IRELAND. *Learning for Life: White Paper on Adult Education* [online]. Dublin: Stationery Office, 2000. Available from < http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1a/c6/5e.pdf> (cited 15th of September 2009).

DERVIN, B. “Strategies for dealing with human information needs: Information or communication?”, *Journal of Broadcasting*, 1976, vol. 20, no. 3, p. 324-351.

DERVIN, B. “An overview of sense-making research: concepts, methods, and results”, *Annual Meeting of the International Communication Association*, Dallas, 1983.

DERVIN, B.; NILAN, M. “Information needs and uses”, *Annual Review of Information Science and Technology,* 1986, vol. 21, p. 3-33.

DERVIN, B. “On studying information seeking methodologically: the omplications of connecting metatheory to method“, *Information Processing and Management*, 1999, vol.25, no. 6, p.727-750.

DONMING, X; HUAIQING, W. “Intelligent agent supported personalization for virtual learning environments”, *Decision Support Systems*, 2006, vol. 42, no. 2, p. 825-843,

DÖRNEY, Z. “Motivation and motivating in the foreing language class-room”, *The modern language journal*, 1994, vol.78, no.3, p. 3-33.

DYCHÉ, J. “Personalization and collaborative filtering”, *The CRM Handbook*,, Boston: Addison-Wesley Information Technology Series, 2002.

EISENBERG, M; LOWE, C. A.; SPITZER, K. L. (2004). *Information literacy: Essential skills for the information age* (2nd ed.). Wetport, CT. Libraries Unlimited, 2004.

ELLIS, D. “Ellis's Model of Information Seeking Behavior”, *Theories of information behaviour*, FISHER, K. E., ERDELEZ, S i MCKECHNIE, L (eds), Information Today, Medford, NJ, 2005, p. 138-143 (ASIST Monograph Series).

EUROPEAN COMMISSION, DG EDUCATION AND TRAINING. “Implementation of the Education and Training 2010 Work Programme, Working Group B: Key Competences”, *Key Competences for Lifelong Learning: a European Reference Framework* [online], 2004. <http://ec.europa.eu/education/policies/2010/doc/basicframe.pdf> (cited 15th of September 2009).

FISHER, K. E.; JULIEN, H. “Information Behaviour”, *Annual Review of Information*

Science, 2009, vol. 43, chapter 7, p. 317-358.

FISCHER, G. *Lifelong Learning - More than Training" in Journal of Interactive Learning Research, 2*000, vol.11, no. 3/4 pp 265-294.

FRIESEN, N. (2002), "Semantic interoperability, communities of practice and the Cancore learning

object metadata profile", Alternative Paper Tracks, *Proceedings of the 11th World Wide Web Conference (WWW 2002)*, held at Hawaii, USA, 2002.

FOSTER-JONES, J.; BEAZLEIGH, H. "Metadata in the changing learning environment: developing skills to achieve the blue skies", *Association for Learning Technology Journal*, 2002, vol. 10, no. 1, p. 52-59.

FORD, N. "Cognitive styles and virtual environments", *Journal of the American Society for Information Science*, 2000, vol. 51, no. 6, p. 543-557.

FULCHER, G.; LOCK, D. "Distance education. The future of library and information services requirements". *Distance Education*, 1999, vol. 20, no. 2, p. 313-329.

GIVEN, L. M. *The social constructions of the "mature student" identity. Effects and implications for academic infomration behaviour*. Londres, Ontario: University of Western Ontario. Diss., 2000.

GLAZIER, J. D.; POWELL, R. *Qualitative research in information management, Englewood, Colorado:* Libraries Unlimited, 1992.

GODOY, C. "Chapter 3. Modeling interest of web users for recommendation: a user profiling approach and trends" In: Richi Nayak, Nikhil Ichalkaranje, and Lakhmi C. Jain (ed.). *Evolution of the Web in artificial intelligence environments,* Berlin: Springer–Verlag, 2008, p. 41-69. (Studies in computational intelligence, 130)

GONZÁLEZ TERUEL, A. *Los estudios de necesidades y usos de la información: fundamentos y perspectivas actuales*. Gijón: Trea, 2005. (Biblioteconomía y Administración Cultural, 123).

GRUBER, T. "A translation approach to portable ontology specifications", *Knowledge acquisition*, 1993, vol. 5, no. 2, p.199-220.

GRUBER, T. "Towards principles for the design of ontologies used for knowledge sharing", *International Journal of Human-Computer Studies*, 1995, vol. 43, p. 907-928.

DOMINGUE, J.; MOTTA, E. "A knowledge-based news server supporting ontology-driven story enrichment and knowledge retrieval", *Knowledge Acquisition, Modeling and Management: 11th European Workshop, EKAW '99 Proceedings*, held at Dagstuhl Castle, Germany, May 26-29, 1999, D. Fensel and R. Studer (eds.). Berlin: Springer, 1999.

QIN, J.; PALING, S. “Converting a controlled vocabulary into an ontology: the case of GEM”, *Information Research* [online], 2001, vol. 6, no. 2. Available from <http://informationr.net/ir/6-2/paper94.html (cited 15th of September 2009)

HARLEY, D.; HENKE, J. “Toward an Effective Understanding of Website Users. Advantages and Pitfalls of Linking Transaction Log Analyses and Online Surveys” [online], *D-Lib Magazine* [online], 2007, vol. 13 no. 3/4. Available from <http://www.dlib.org/dlib/march07/harley/03harley.html> (cited 15th of September 2009).

HARRISON, R.; REEVE, F. HANSON, A. (eds.). “Introduction. Perspectives on learning”, *Supporting Life Long Learning Volume I: Perspectives on learning*, Milton Keynes: The Open University, 2002.

HEFLIN, J. (ed.) “OWL Web Ontology Language Use Cases and Requirements. W3C” [online] *W3C Recommendation 10 February 2004.* Available from <http://www.w3.org/TR/webont-req/> (cited 15th of September 2009).

HEPWORTH, M. “Knowledge of information behaviour and its relevance to the design of people-centred information products and services”, *Journal of Documentation*, 2007, vol. 63, no. 1, p. 33-56.

HERLOCKER, J.L.; KONSTAN, J. A.; TERVEEN, L. G.; RIEDL, J. T. “Evaluating collaborative filtering recommender systems”, *ACM Transactions on Information Systems*, 2004, vol. 22, no. 1, p. 5-53.

IIOVEN, M.; HUOTARI, M. “The impact of trust on the practice of knowledge management”, *Proceedings of the 63rd ASIS Annual Meeting,* held at Chicago 11-16 November. Softbound: 2000, vol. 37, p.421-429.

INTERNATIONAL INFORMATION AND COMMUNICATION TECHNOLOGIES (ICT) LITERACY PANEL. *Digital transformation: A framework for ICT Literacy* [online]. Princeton, NJ: Educational Testing Services (ETS), 2002. Available from <http://www.ets.org/research/ictliteracy/ictreport.pdf> (cited 15th of September 2009).

JONES, W. “Personal Information Management”, *Annual Review of Information Science*, 2007, vol. 10, p. 453-505.

JOHNSTON, B.; WEBBER, S. “Information literacy in higher education: a review and case study”, 2003, *Studies in Higher Education*, vol. 28, no. 3, p. 335-352.

KASANOFF, B. *Making it personal*, Cambridge, MA: Perseus Publishing, 2001.

KATZ, J. E.; RICE, R. E. “Social consequences of internet use”, Ac*cess, involvement and interaction*. Cambridge, MA: The MIT Press, 2002.

KARI, J. “Conceptualizing the personal outcomes of information” [online], *Information Research*, 2007, vol. 12, no. 2. Available from <http://InformationR.net/ir/12-

2/paper292.html> (cited 15th of September 2009).

KIM, K. S."Information seeking on the Web: effects of user and task variables" *Library and Information Science Research,* 2001, vol. 23, no. 3, p. 233-255.

KLEINBERG, J.; SANDLER, M. "Using mixture models for collaborative filtering"*, Journal of Computer and System Sciences*, 2008, vol. 74, no. 1, p. 49-69.

KNOWLES, M.S.; Holton, E.; Swanson, R.A. *The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development*. Houston, TX: Gulf Publishing Company. for Studies in Education., 1998.

KRIKELAS, J. "Information-seeking behaviour: patterns and concepts", *Drexel Library Quaterly*, 1983, vol. 19, p.5-20.

KRIPPENDORFF, K. *Content Analysis. An Introduction to its methodology*. Londres, Los Angeles: Sage Publications, 1980.

KUHLTHAU, C.C. "Inside the search process: information seeking from the users perspective"*, Journal of the American Socety for Information Science*, 1991, vol. 42, no. 5, p. 361-371.

KUHLTHAU, C.C. "Learning in digital libraries. An information search process approach", *Library Trends*, 1997, vol. 45, no. 4, p. 708-725.

LAMBE, P. *Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness*. Oxford: Chandos Publishing, 2007.

LAW, E. L.; KLOBUCAR, T.; PIPAN, M. "Analysis of user attitude and behaviour in evaluating a personalized search engine", In *Proceedings of the 13th Eurpoean Conference on Cognitive Ergonomics (ECCE): Trust and Control in Complex Socio-Technical Systems, held at* Zurich, Switzerland, September 20 – 22 2006. New York, NY: ACM, 2006.

LECKIE, G. J. "General model of the information seeking professionals", Fisher, KE, Erdelez, S i McKechnie, L (Eds), *Theories of information behaviour*, Medford, NJ: Information Today, 2005. p. 158-164 (ASIST Monograph Series).

LÉVEILLÉ, Y. *La recherche d'information à l'école secondaire: l'enseignant et le bibliothécaire, partneaires de l'éléve* [online], 1997. Available from: <http://pages.infinit.net/formanet/cs/chap2.html [Cited 25th of September 2009].

LINDEN, G.; SMITH, B. YORK, J. "Amazon.com Recommendations. Item-to-Item Collaborative Filtering", *Internet Computing, IEEE,* vol. 7, no. 1, Jan/Feb 2003, p. 76 – 80.

LLOYD, A. "Information literacy: The Meta-Competency of the Knowledge Economy? An exploratory paper", *Journal of Librarianship and Information Science*, 2003, vol.35,

no. 2, p. 87-92.

LLOYD, A.; SOMERVILLE, M. “Working information”, *Journal of Workplace Learning*, 2006, vol. 18, no. 3, p. 186-198.

LOM. *Draft Standard for Learning Object Metadata* [online], 2002. Available at: <http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf> [Cited 25th of September 2009].

LYNCH, C. “Digital library opportunities”, Journal of Academic Librarianship, 2003, vol. 29, no. 5, p. 286-9.

MCKECHNIE, P.J., "A model of information practices in accounts of everyday life information seeking", Journal of Documentation, 2003, vol. 59, p.19-40.

MOR, E.; MINGUILLÓN, J. “E-learning personalization based on itineraries and long-term navigational behavior”, *Proceedings of the 14th International World Wide Web Conference*, Alternate Track Papers and Posters, held at New York, EUA, 2004, p. 264-265.

NAYAK, R. JAIN, L. C. “Chapter 2. Innovations in Web Applications Using Artificial Intelligence Paradigms” In: Richi Nayak, Nikhil Ichalkaranje, and Lakhmi C. Jain (editors). *Evolution of the Web in artificial intelligence environments*. *Studies in computational intelligence*, volume 130. Berlin: Springer–Verlag, 2008, p. 17-41.

NEVEN, F.; DUVAL, E. “Reusable learning objects: a survey of LOM-based repositories”, Palmer, J.. "Scientists and information: II. Personal factors in information behaviour", *Journal of Documentation*, 2002, vol. 47, no. 3, p. 254-275.

NEUHOLD, E.; NIEDERÉE, C.; STEWART, A. “Personalization in Digital Libraries: an extended view”, *Proceedings of the 6th International Conference on Asian Digital Libraries (ICADL 2003)*, LNCS 2911, p. 1–16 held in 8 - 11 December 2003, Kuala Lumpur, Malaysia

ROWLAND, I.; NICHOLAS, D. “Understanding information behaviour: how do students and faculty find books”, *The Journal of Academic Librarianship*, 2008, vol. 34, no. 1, p.3-15.

NOWÉ, K.; MACEVICIUTE, E.; WILSON, T. D. “Tensions and contradictions in the information behaviour of Board members of a voluntary organization"[online], *Information Research*, 2008, vol. 13, no. 4. Available from <http://InformationR.net/ir/13-4/paper363.html>

NATIONAL INFORMATION STANDARDS ORGANIZATION (NISO), *Understanding Metadata, NISO Press*, Bethesda, USA, 2004.

NIELSEN, J. “Personalization is over-rated” [online], *Alertbox*. Available from <http://www.useit.com/alertbox/981004.html> [Cited 25th of September 2009].

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD) *Learning to bridge the digital divide,* Paris: OECD, 2000.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD) *E-learning in tertiary education. Where do we stand?,* Paris: OECD, 2005.

OPEN E-LEARNING CONTENT OBSERVATORY SERVICES (OLCOS). "Open Educational Practices and Resources. OLCOS Roadmap 2012" [online], G. Geser (ed.), Salzburg Research, EduMedia Group, Salzburg, 2007.Available from <http://www.olcos.org>

PETTIGREW, K. E., FIDEL, R.; BRUCE, H. "Conceptual frameworks in information behavior". *Annual Review of Information Science and Technology,* 2001, vol. 35, p. 43-78.

PÉREZ-MONTORO GUTIÉRREZ, M. *Arquitectura de la Información en entornos web*. Gijón: Trea. 2010. Forthcoming.

RAY, K.; DAY, J." Student attitudes towards electronic resources" [online], *Information research*, 1998, vol. 4, no. 2. Available from: <http://informationr.net/ir/4-2/paper54.html> [Cited 9 October 2003].

RECKER, M. M.; WILEY, D. A. "A non-authoritative educational metadata ontology for filtering and recommending learning objects", *Journal of interactive learning environments*, 2001, vol. 9, no. 3, p. 255-271.

RICHARDS, T.; RICHARDS, L. "Using computers in qualitative research", *Collecting and Interpreting Qualitative Materials, 1998,*. Ed. Norman K. Denzin and Yvonna S. Lincoln. London Sage Publications, 1998, p.211-245

RIECKEN, D. "Personalized views of personalization", Communications of the ACM, 2000, vol. 43, no. 8, p. 27-8.

SANZ CASADO, E. *Manual de estudios de usuarios*, Madrid: Fundación Germán Sánchez Ruipérez. 2004.

SARADEVIC, T.; KANTOR, P. B. "Studying the value of library and information services. Part I. Establishing a theoretical framework", *Journal of the American Society for Infomration Science*, 1997, vol. 48, p. 527-542.

SAVOLAINEN, R. "Everyday life information seeking: approaching information seeking in the context of way of life", *Library and Information Science Research*, 1995, vol. 17, no. 2, p. 259-84.

SAVOLAINEN, R. "Enthusiastic, realistic and critical. Discourses of Internet use in the context of everyday life information seeking "[online], Information Research, 2004, vol. 10, no. 1, Available from <http://InformationR.net/ir/10-1/paper198.html> [Cited 25th of September 2009].

SAVOLAINEN, R. ”Everyday life information seeking”, Fisher, KE, Erdelez, S i McKechnie, L (Eds), *Theories of information behaviour*, 2005, Information Today, Medford, NJ, 143-8 (ASIST Monograph Series).

SCONUL Society of College, National and University Libraries. The *Seven Pillars of Information Literacy* [online]. *A*vailable from <http://www.*sconul*.ac.uk/groups/information_literacy/seven_pillars.html>

SCHREIBER, T.; HARBO, K. “Information literacy and personal knowledge management” [online], *Nord I&D, Knowledge and Change*, 12th Nordic Conference on Information and Documentation, Aalborg, Denmark, September 1-3, 2004. Available from <http://www2.db.dk/NIOD/schreiber-harbo.pdf>

SCHONBERG, E.; COFINO, T.; HOCH, R.; PDLASECK, M.; SPRARAGEN, S. L. “Measuring success”, *Communications of the ACM*, 2000, vol. 43 no. 8, p. 53-7.

SHETH, A.; RAMAKRISHNAN, C.; THOMAS, C. “Semantics for the Semantic Web: the implicit, the formal and the powerful”, *International Journal on Semantic Web & Information Systems*, 2005, vol. 1 no. 1, p. 1-18.

SEVERANCE, C.; HARDIN, J.; WHYTE, A. “The coming functionality mashup in personal learning environments, Interactive Learning Environments Journal, vol. 16, no. 1, 2008.

SINHA, R.; SWEARINGEN, K. “Comparing recommendations made by online systems and friends”, Proceedings of the Second DELOS Network of Excellence Workshop on Personalization and Recommender Systems in Digital Libraries, 2001. Available from <http://www.ercim.org/publication/ws-proceedings/DelNoe02/RashmiSinha.pdf >

SICILIA, M. A.; GARCÍA, E. “On the convergence of formal ontologies and standardized e-learning”, *Journal of Distance Education Technologies*, vol. 3, no. 12, p. 12-28

SOLLUND, A. “Personalisation through the use of personal profile”, *ePerSpace*, Fornebu, Norway, 2007.

SPINK, A; COLE, C. “Introduction”, *Journal of the American Society for Information Science and Technology*, 2004, vol. 55, no. 8, p.657-659.

SUTHERLAND, P. *Adult learning: a Reader*. London: Routledge, 1999.

TALJA, S. *“The Domain Analytic Approach to Scholars'Information Practices”, Theories of information behaviour*, Fisher, KE, Erdelez, S i McKechnie, L (Eds), *Theories of information behaviour*, Information Today, Medford, NJ, 2005, p. 123-128 (ASIST Monograph Series).

TAYLOR, R.S. "Information use environments". In Dervin, B. and M.J. Voight (Eds.). *Progress in Communication Sciences*, vol 10. Norwood, NJ: Ablex Publishing Corp., 1991, p. 217-255.

TOLEDANO, R. "Information literacy and knowledge management: preparations for an arranged marriage", *Libri*, 2008, vol. 58, no. 3, p. 155-171.

THORSTEINSDOTTIR, G. *The information seeking behaviour of distance students*. Doctoral thesis, Borås: The university college of Borås, Valfrid. 2005.

UNESCO *et al. Goals, objectives and participant responsibilities. Meeting of experts on information literacy*. 2003 Availavle from <http://www.nclis.gov/libinter/infolitconf&meet/goals-objectives-participant_responsibilities.html>

VIRKUS, S. "*Development of information-related competencies in European ODL institutions: senior managers' view*". *New library world*, 2006, vol. 107, no. 11/12, p.467 – 481.

W3C. "W3C Personalization Roadmap: Ubiquitous Web Integration of AccessForAll", *Working Group Note 09 April 2009 1.0,* 2009. Available from <http://www.w3.org/TR/2009/NOTE-UWA-personalizationRoadmap-20090409/>

WALDMAN, M. "Freshmen's Use of Library Electronic Resources and Self-Efficacy, *Information Research*, 2003, vol. 8, no. 2. Available from: <http://informationr.net/ir/8-2/paper150.html> (Cited 16th September 2009)

WEBBER, S., BOON, S.; JOHNSTON, B. "A comparison of UK academics' conceptions of information literacy in two disciplines: English and Marketing", *Library and information research,* 2005, vol. 29, no. 93, p. 4-15.

WILKINSON, M. A. "Information sources used by lawyers in problem solving: An empirical exploration", *Library & Information Science Research*, 2002, vol. 23, no. 3, p. 257-276.

WILSON, T. D. "On user studies and information needs", *Journal of Documentation*, 1991, vol. 37, 1, p.3-15.

WILSON, T. D. "Models in information behaviour research", *Journal of Documentation*, 1999, vol. 55, no. 3, p. 249-270.

WILSON, T. D. "Human information behaviour", *Information Science*, 2000, vol. 3, no. 2, p. 49-55.

WILSON, T. D. "Revisiting user studies and information needs", *Journal of Documentation*, 2006, vol. 62, no. 6, p. 680-684.

YOUNG, R. M.; BARNARD, P. "The use of scenarios in human-computer interaction

research: Turbocharging the tortoise of cumulative science*"*, In J.M. Carroll & P. Tanner (Eds.), *Proceedings of the Human Factors in Computing Systems and Graphics Interface,* held at Toronto, Canada, 1987. New York: ACM, 1987. p. 291-296

Annex

Paper I

FERRAN, N.; MOR, E.; MINGUILLÓN, J. “Towards personalization in digital libraries through ontologies”, *Library Management Journal*, 2005, vol. 25, no. 4/5, p. 206-217.
ISSN 0143-5124

Afiliació dels autors:
Núria Ferran Ferrer
Information and Communication Sciences Studies, Universitat Oberta de Catalunya
Dr. Enric Mor
Computer Science, Multimedia and Telecommunications Studies, Universitat Oberta de Catalunya
Dr. Julià Minguillón Alfonso
Computer Science, Multimedia and Telecommunications Studies, Universitat Oberta de Catalunya

Resum en català:
A l'article “Towards personalization in digital libraries through ontologies” s'identifiquen els elements que determinen les funcionalitats bàsiques d'un sistema de personalització d'una biblioteca digital d'un entorn virtual d'aprenentatge. Els elements identificats són el perfil d'usuari, que inclou l'històric de navegació i les preferències de l'usuari i després la informació dels comportaments de navegació de la biblioteca digital.

The Emerald Research Register for this journal is available at
www.emeraldinsight.com/researchregister

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0143-5124.htm

Towards personalization in digital libraries through ontologies

Núria Ferran
Digital Library, Universitat Oberta de Catalunya (UOC), Barcelona, Spain, and
Enric Mor and Julià Minguillón
Computer Science and Multimedia Studies, Universitat Oberta de Catalunya (UOC), Barcelona, Spain

Abstract

Purpose – To describe a browsing and searching personalization system for digital libraries based on the use of ontologies for describing the relationships between all the elements which take part in a digital library scenario of use.

Design/methodology/approach – Identification of all the desired functionalities and requirements that are necessary to fully integrate the use of a digital library in an e-learning environment, and the basic elements that are used to build the ontology that describes such scenario.

Findings – The elements that determine the functionalities of the desired personalization system: first, the user's profile, including navigational history and user preferences; and second, the information collected from the navigational behavior of the digital library users.

Research limitations/implications – The ontology is not complete. In fact, the ontology in itself will evolve with the new apparition of desired functionalities and requirements of the personalization system.

Practical implications – Such a personalization system will be very helpful to the users of a digital library to improve their experience of use.

Originality/value – The use of ontologies promotes the integration of new services into existing ones, and the interoperability with other systems through the appropriate semantic web services. New system functionalities and requirements can be added by including the appropriate description into the ontology framework that defines the digital library scenario of use.

Keywords Digital libraries, Information searches, Personal needs, Customization, Distance learning

Paper type Research paper

Introduction

Distance education is becoming one of the most attractive methods for incorporating all kinds of people into higher and university degree education levels, moving towards a "blended" technology approach deploying multiple technologies. The introduction of new technologies of information and communications with the intensive use of e-learning environments, as a virtual campus, for example, allows students to break through the barriers of space and time, and to design their own lifelong curricula, adapting it to their particular necessities and preferences, according to their possibilities as students, changing the usual way of both teaching and learning (Jonassen *et al.*, 1995), setting up the foundations of e-learning environments (Rosenberg, 2001).

The students of an e-learning environment have access to a predetermined repository of learning resources, which are part of the learning process designed by the

Library Management
Vol. 26 No. 4/5, 2005
pp. 206-217
© Emerald Group Publishing Limited
0143-5124
DOI 10.1108/01435120510596062

team of instructional designers and teachers for each course. But usually, these students might need more additional learning resources and documents to successfully follow the recommended learning itineraries, so there is the necessity of providing mechanisms for accessing such resources in a structured manner (Sicilia and García, 2005). On the other hand, researchers and teachers in an e-learning environment are also usual "customers" of the services offered by an academic library, although under a completely different approach. These kinds of users have different experiences of use of the digital library than students, as their goals are clearly different, in both content and context aspects. All this richness needs to be captured somehow in order to better understand the way users of a digital library perform their navigation. Digital libraries must evolve in order to be proactive, more responsive to possible changes and to include new services such as personalization in order to increase user satisfaction and fidelity.

Therefore, in an ideal scenario, the digital library should adapt to the specific characteristics of each user profile, but also to the particular necessities and preferences of each user, combining both user and profile level personalization capabilities. Personalization is one of the key factors which are directly related to user satisfaction (Riecken, 2000) and, therefore, linked to the failure or success of the performed activity, although it must be carefully introduced (Nielsen, 1998). Personalization has been shown useful in several areas such as e-commerce (Kasanoff, 2001), business-to-business companies (Colkin, 2001), and obviously reproduced in other environments such e-learning (Mor and Minguillón, 2004), for example. Regarding the library management field, there exist remarkable approaches such as the recommendation system for electronic journals of MyLibrary from Los Alamos National Laboratory 2.4 (http://lib-www.lanl.gov/) or MyOpenLibrary from the Open University (http://library.open.ac.uk/index.html); such personalization librarian initiatives are showing successful results as appears at the research being done about the user satisfaction (Sinha and Swearingen, 2001).

In order to build such personalization system, several multidisciplinary aspects must be addressed: first, there are cognitive and behavioral aspects (Ford and Ford, 1993) that determine the way users perform searches and examine the obtained results. The "I'm Feeling Lucky" button in the Google search engine home page is a good example of such fact. Second, personalization issues must be addressed from a user-centered point of view, under the approach of human computer interaction, as it is well known (Schonberg *et al.*, 2000) that most personalization systems fail, not because of the personalization system in itself, but in the interaction with the user and the way recommendations are presented. Third, there are technological and knowledge engineering aspects related to the way all this information is structured for both updating and querying purposes. In this paper we describe the set of desired functionalities and requirements of an ideal scenario for a digital library which includes personalization capabilities by means of ontologies. The use of ontologies for describing the possible scenarios of use in a digital library brings the possibility of predicting user requirements in advance and to offer personalized services ahead of expressed need. Ontologies are built using other sub-ontologies which describe the basic elements of the personalization system: users, digital resources, actions, navigational profiles, etc. This diversity of elements suggests that there is a need for further focus on the interoperability of objects which in turn requires well developed

ontologies to describe the properties of both objects and individuals and the relationships between them (Brophy, 2004).

This paper is structured as follows: the second section describes the basic functionalities of a digital library and the specific case of the UOC digital library, and the fundamental concepts of ontologies used in this paper. The third section defines the requirements and functionalities of a personalized digital library integrated in a virtual e-learning environment. All the elements defined as part of the ontology which gives support to the personalization system are described in the fourth section. Finally, conclusions and future research directions are outlined in the fifth section.

Digital libraries and the case of the UOC

As stated previously, one of the most important resources for supporting users in a distance e-learning environment is the possibility of accessing to a digital library, which allows the users to collect and organize the necessary information for achieving their particular goals. Furthermore, the search of information can be a learning but also an assessment activity by itself, so it is important to ensure and facilitate a proper use of the library.

There are several terms being used interchangeably when we approach the concept of a library with digitized data and accessible remotely. Among these we can find hybrid library, digital library and virtual library. An informal definition of a digital library is "a managed collection of information, with associated services, where the information is stored in digital formats and accessible over a network" (Arms, 2002). The hybrid library provides electronic information sources too but also paper-based information. The hybrid should be considered as a model by itself not as a transitional phase from a conventional library to a digital one (Brophy, 2001). And finally the virtual library has been defined as the concept of a "remote access to the contents and services of libraries and other information resources, combining an on-site collection of current and heavily used materials in both print and electronic form, with an electronic network which provides access to, and delivery from, external worldwide library and commercial information and knowledge sources" (Gapen, 1993).

Nowadays, we are finding new types of libraries coming up from long-term personal digital libraries, as well as digital libraries that serve specific organizations, educational needs, and cultural heritage and that vary in their reliability, authority and quality. Besides, the collections are becoming more heterogeneous in terms of their creators, content, media, and communities served. In addition, the user communities are becoming heterogeneous in terms of their interests, backgrounds, and skill levels, ranging from novices to experts in a specific subject area (Callan and Smeaton, 2003). This growing diversity has changed the initial focus of providing access to digital content and transforming the traditional services into digital ones to a new handicap where the next generation of libraries should be more proactive offering personalized information to their users taking in consideration each person individually (his or her goals, interests, level of education, etc.).

While data and information are captured and represented in various digital formats, and rapidly proliferating, the techniques for accessing data and information are rudimentary and imprecise, mostly based on simple keyword indexes, relational queries, and/or low-level image or audio features (i.e., research results of the 1970s and 1980s). In the current context of explosive availability of data, there is a need for a knowledge discovery approach, based on both top-down knowledge creation (e.g.,

ontologies, subject headings, user modeling) and bottom-up automated knowledge extraction (e.g., data mining, text mining, web mining), promises to help transfer digital library from an institution of data and information to an institution of knowledge (Chen, 2003).

The UOC virtual library

The UOC Library was born in 1995 as a virtual academic library to support a virtual e-learning university model and, since then, provides online services and information resources both print and digital, owned by the library or by other libraries. Users can access the library from any computer and do not need to move in order to get any information resource. All authorized users are able to use remotely the exclusive content of the library as subscribed databases like the Electronic Management Research Library Database (Emerald) or the Web of Knowledge from ISI and benefit from the services of the library such as the Selective Diffusion of Information, Bibliographic Searches or Loans, all of them performed through the web.

The UOC virtual library can be accessed in different ways. One of them is from the virtual campus where the user finds the whole content and services of the library. But the main focused (i.e., guided) entrance to the library can be found in the campus's virtual classrooms where teachers and librarians bring a selection of the most interesting resources for every subject, for instance the learning material, recommended bibliography where each book is linked with the loan form, electronic articles, self-assessment exercises, a selection of internet resources, databases and electronic journals, e-books, exams from previous years, etc. This has been the first step for tailoring information for a very well defined community.

The main objective of the library is to provide the students, but also lecturers, researchers and management staff, access to the information relevant for the fulfillment of their basic functions: learning, teaching, research and management. Several user profiles can be identified: undergraduate student, PhD student, teacher, learning process manager, among others. Each profile can be partially identified by the tasks related to the digital library that it performs. For example, students usually browse the digital library looking for documents related to activities such as exercises, exams and recommended articles and in very specific periods of time, when the delivery date of the exercise is due. On the other hand, teachers can navigate among the content of the library in order to mentor a student in doing his or her homework or providing content to the digital library associated to the virtual classroom. A final example could be the researchers who usually perform more focused searches during a larger period of time. Each of these users may have common goals in certain times but their knowledge, tasks, social activities and preferences are totally different.

Ontologies and the Semantic Web

Therefore, it is necessary to build a complete and complex structure for describing all the richness of the possible scenarios of use of the digital library and the relationships which can be established among all the participants. This can be achieved by means of ontologies and the use of the Semantic Web services in their appropriate forms (Sheth *et al.*, 2005).

An ontology is, taking the meaning adopted by the semantic web community, a formal description of a possible scenario or context; that is, what "exists" is what can be represented by an ontology. Formally, an ontology is the statement of a logical

theory, but by "formal description" we also mean that it can be automatically queried and updated, as the main users of ontologies are (or should be) computers, not humans, in order to explicitly represent the objects, concepts and other entities that are assumed to exist in some context, altogether with the relationships that hold among them, although ontologies must be also human-readable. "Ontologies and taxonomies are, in functional terms, often used as synonyms. Computer scientists call hierarchies of structured vocabularies *ontologies* and librarians deploy the term *taxonomy*" (Adams, 2002). Nevertheless, ontologies also include a set of semantic rules which are used to infer knowledge from a structured hierarchy of information, giving to the complete structure a semantic meaning, not only syntactic (Gruber, 1995).

Requiriments for an integrated, personalized digital library

The web has become a very common tool for information browsing and searching, and the success of search engines such as Google or A9, for example, has facilitated the diffusion and access to repositories of digital documents. Despite that, one of the main problems of such search engines is that the generated results are not always of interest for the users performing the search, as these engines use a generalist approach based on several criteria which might not match the criteria of a specific user. On the other hand, several e-commerce web-based services, such as Amazon or e-Bay, for example, also provide browsing and searching services, but focused on categories. Both approaches can be combined to facilitate the way users browse the contents in search of information.

Several techniques are used for guidance and for providing recommendations to users; among others, collaborative filtering (Herlocker *et al.*, 2004) is one of the most successful ones. Briefly, collaborative filtering is selecting content based on the preferences of people with similar interests, basically by pooling and ranking informed opinions (or experiences of use) on any particular topic. That is to say, an automatic system collects information about user actions (explicit, such as voting or answering a question; or implicit, such as noticing which offered links are visited and which are not, and how much time) and determines the relative importance of each content by weighting all the collected information among the large amount of users.

Both navigational techniques are also valid in a digital library scenario of use: simple searches starting from a single search term or advanced searches using multiple criteria, but also a recommendation system based on guided navigation through an ordered set of categories. The basic idea of this paper is that the efforts for finding a useful piece of information in a digital library carried out by an individual can be stored in a structured way and then shared for future users with similar necessities. Furthermore, if such information searching and browsing combines several web-based resources with different approaches (access mechanisms, query languages and interfaces, and so on), it is important to describe a common strategy (Sadeh and Walker, 2003) for minimizing the necessary efforts to fight against duplication (thus inconsistency) and source diversity.

Identification of personalization system functionalities

Two elements determine the functionalities of the desired personalization system: first, the user's profile, including navigational history and user preferences; and second, the information collected from the navigational behavior of the digital library users. User profile should include all the information relevant to user: personal information, which

can be publicly made available by each user in order to facilitate the discovery of similar interests; and navigational history and behavior records, which will be used altogether with the personal information by the personalization system to build the set of recommendations that will help each user in browsing and searching the digital library. This information should help the user to improve his or her searches, by obtaining additional information when searching or browsing. It is remarkable to say that this information has been validated by the ontology, and that is not biased by any non-academic purpose of use (such as commercial-supported recommendations in Google or Amazon, for example). Table I shows the basic user profile attributes which are used to build the user model. For each attribute, it is shown whether it can be obtained or it must me given by the user, the relative importance for navigation and the recommendation system, and the user profiles which is more related to.

For instance, the socio-demographical attribute may include information about the languages spoken or understood by the user, information that can be used as an additional filtering when browsing or searching for documents. This attribute has a medium importance with respect to personalization of navigational actions, and a low importance for the recommendation system. On the other hand, other attributes which express user preferences and interests have a high importance for both navigational and recommendation system actions. Other information about profiles such as information about factual/behavioral issues (Adomavicius and Tuzhilin, 1999) can be also included at this level.

Two different behavior types can be identified, depending on the users' navigation, exploratory navigation and goal-oriented navigation. The exploratory navigation can be mainly oriented to obtain a general vision of the available resources in the library. Depending on user profile, the exploratory navigation would have different implicit intentions. In the case of goal-oriented navigation, it is usually considered that the user is looking for a resource. These searches can be classified in different use cases. For example, in the situation of searching for an author, if the user is a student, the

Attributes	Is it explicit?	Navigational support value level	Library recommendation level	Profiles concerned
Academic Register, Actual enrollment	No	High	High	Student
Academic Register, Previous enrollment	No	Low	High	Student
Navigational history	No	High	Medium	All
Navigational behavior and strategies	No	High	Medium	All
Socio-demographical profile, General background	No	Medium	Low	All
Socio-demographical profile, Academic background	No	Low	Medium	Student
Knowledge area	Yes	High	High	Teacher, Consultant, Researcher
Interests	Yes	High	High	All
Preferences	Yes	High	High	All

Table I.
Basic user profile attributes

recommendations associated to search results should be oriented to the area of the course subjects, taking into account the navigation of other students and also the recommendations of the teachers. If the user is a researcher, recommendations should be oriented by different criteria depending on the searches that have been carried out by other investigators, or to the magazines, books and conferences where the searched author had published, understanding that the same magazine, conference or book might contain other interesting resources. Recommendations are generated using the knowledge extracted from the searching and browsing profiles of users with similar interests, knowledge integrated in the ontology such as course bibliography, or by following citations of similar documents, for example. Regarding the sources of information, using the library for accessing selected free Internet resources will be of particular interest, because the recommendation system ensures the users receive the opinions of a large set of experts (that is, the collaborative filtering system), therefore giving authenticity to such electronic information sources. On the other hand, regarding the library exclusive content, the user will get into the external databases (commercial) or internal databases (such as digital repositories or catalogues) in a transparent manner. Table II shows all the current information sources present in the digital library, and the basic actions that can be taken by the library users.

Navigation	Actions	Influence
Catalogue		
Searching	Search physic and analogical documents (books, VHS, DVD, etc.) into the OPAC	Low
Browsing	Navigate through author, subject, keyword, title, title course	Medium
Browsing	Look into the bibliographic register	Medium
Browsing	Look into the abstract and table of contents (for each document the DL scans the front cover, back cover, index)	Medium
Command	Loan command	High
Digital collection		
Browsing	Navigate among the thematic classification nodes	Medium
Searching	Search inside and electronic database	Medium
Command	Download, print, saving searches, sending searches by e-mail, create a Table of Contents (TOC) alert, add a journal to favorites, etc.	High
Subscription services		
Command	Subscription to the News Services (for each study the librarians create a weekly newsletter with a news selection about the subject of the study. Users received it as an e-mail)	Medium
Command	Subscription to paper journal TOC	High
Command	Subscription to SDI (Selective Diffusion of Information)	High
Library classroom		
Browsing	Look into the recommended bibliography	Medium
Browsing	Look into the FAQs	Medium
Browsing	Look into the didactic material	Medium
Browsing	Look into exercises of other semesters	Medium

Table II.
Information sources and basic user actions in the digital library

There are more possible actions that are currently being performed in the context of the digital library scenario of use but that are not integrated yet. For instance, teachers usually recommend a basic bibliography for any subject, which is supposed to be used by students for solving the course exercises. The use of an ontology could be also interesting to incorporate such new functionalities into the existing digital library, by describing the relationships between elements. In plain words: if a teacher defines one or more books as recommended bibliography for a given subject, students enrolled in such subject should be aware of those books when performing searches related to the subject. Discovering these new functionalities requires from the study of the current user behavior (taking into account the different user profiles) in the virtual e-learning environment, by means of classical HCI methodologies (Dix *et al.*, 1998), involving users in the design of the new digital library. Then, these functionalities and requirements can be described as rules ranging from simple to complex statements within the ontology framework.

For library managers, the creation of an ontology will help them to construct tailored libraries for each subject. Every library is built on the explicit recommendations from a teacher, but in an unstructured manner. With an ontology, these specialized libraries could be built from the use that previous students gave to that resources and new information could be added from the use of the library by experts. The ontology itself is composed of sub-ontologies. Thus, the digital resources, which are catalogued using the Dublin Core metadata or MARC format among other standards can be extended by means of an ontology to include additional information, automatically or in explicit way, such as their usage, user ratings and any other useful information that users might consider, such as summaries, enhanced taxonomies and keywords, for example.

Privacy issues

A very important aspect that cannot be ignored is the fact that users are always under control, in the sense that all taken actions are monitored and registered. This might seem a very invasive setup which harms user privacy and, therefore, undesirable. Nevertheless, there are several remarkable facts that need to be clarified:

- users know in advance that, in a virtual e-learning environment (or any other web based environment), all actions are logged;
- the recommendation system must be designed in a non-intrusive manner and be user-friendly, including the possibility of disconnecting it or minimizing its participation in the browsing or searching activities; and
- the participation of each individual user in the final recommendation system is completely anonymous.

Finally, it is also important to remark that the collected information is only used with personalization purposes, and it is not meant for commercial reasons, and that the library (a non-profit organization) will use the data rationally and in a transparent way. As usual, a tradeoff between personalization and privacy must be established. The more information the user reveals, the more personalized services he or she obtains.

On the other hand, the browsing and searching history of each user is part of his or her private profile, and only the user can modify (delete) his or her history records in order to update his or her preferences and navigational profile, according to the directions given by the personalization system. Some parts of this profile can be made

publicly available in order to create strong links with other users sharing the same navigational interests. There is evidence that people are sometimes eager to be identified and become part of a community (Lynch, 2003), and the use of ontologies could be one of the strengths of digital libraries for pursuing such purpose.

Elements of a digital library ontology

First, it is important to clarify that we are not building an ontology for describing the contents of a digital library, but an ontology for describing the way users browse and search such contents, with the aim of building a personalization system based on accurate recommendations. Therefore, more than building a low-level ontology for describing a particular concept, we are trying to describe a complex scenario of use. Table III describes the basic steps (Denny, 2002) that should be followed for describing the richness of such complex scenario. As usual, these steps overlap and must be taken in a recursive way: depending on the results of the evaluation undertaken in the fifth step, several definitions in the second and next steps might be modified. In fact, the ontology in itself will evolve with the new apparition of desired functionalities and requirements.

Acquiring domain knowledge and organizing the ontology, are where more efforts must be made to ensure a complete representation of the digital library scenario of use. As described above, Tables I and II outline the basic attributes and actions in the

Step	Description	Resources/participants
Acquire domain knowledge	Assemble appropriate information Define all terms used in the domain to describe elements Check formality and consistency	Expertise from library managers, computer scientists, usability experts
Organize the ontology	Design overall conceptual structure Identify domain's principal concrete concepts and properties Identify relationships among such concepts Minimize possible overlaps and inconsistencies Define actions performed between concepts	Definitions: user profiles digital resources navigational profiles learning activities Personalization system functionalities
Elaborate the ontology	Add all the concepts, relations and individuals to achieve a necessary level of detail Reuse information from the current environment as much as possible	Instances of previously defined concepts
Consistency checking	Reconcile syntactic, logical and semantic inconsistencies Involve automatic classification for defining new concepts and class relationships	Ontology logging
Validate the ontology	Perform a final verification Commit the ontology in a real operative scenario of use Compare the obtained results with the desired functionalities Detect all the possible mistakes or misuses of the real scenario	User tests Data analysis Ontology mining

Table III.
Basic steps for designing ontology

current digital library and e-learning environments, which should be used as basic pieces for building the ontology.

Organization of the digital library ontology

Therefore, the ontology is built from the intersection of the elements in Tables I and II, by describing all the interesting relationships for recommendation purposes. This is carried out by defining small micro-scenarios which reveal typical uses of the digital library and their impact on user profile and navigational history. For instance, suppose the following micro-scenario: John is a researcher working in his PhD in machine learning, as his profile says. He is working with Mary, his advisor, who got her PhD in the same field. Mary is also the advisor of other two students, Peter and Ann, which are also working in machine learning related subjects. When students search for scientific papers, the results are sorted depending on papers' relative importance, according to the number of times each paper has been downloaded by other researchers (giving more importance to the other students which work with Ann) but also on whether such paper has been reviewed or not by any senior researcher (giving more importance to Ann's opinion).With the appropriate tools and ontology definition languages, such statement can be translated into one or more rules that feed the recommendation system with the actions taken by all the users in the digital library.

It is remarkable that the use of ontologies can be also extended to implement and transfer the concept of user profile and user navigational behavior to other digital libraries and databases, so when a digital library user leaves one service to connect into another one, the user profile (including preferences and navigational behavior) can be transferred from one database to another through the appropriate semantic web services, because all databases share a common domain of discourse that can be interpreted further by rules of inference and application logic.

Regarding implementation issues, ontologies are usually described by means of one or more descriptive languages based on XML (W3C, 2004a). Basically, RDF (W3C, 2004b) is used for describing resources, while DAML + OIL (W3C, 2004b) which is currently being evolved into the Web Ontology Language (OWL) standard (W3C, 2004c), is becoming the standard for describing ontologies and accessing resources through the web. The use of XML and description language standards ensures the interconnectivity with other existing ontologies and the possibility of upgrading the ontology for new requirements and functionalities. The widespread of XML for describing data (but also information and knowledge, with the help of ontologies) has made possible the apparition of new applications into classical areas of knowledge.

Conclusions

In this paper we have described the requisites of a personalization system which uses all information relevant to the process of searching and browsing a digital library to build a complete navigational profile for each user and its semantic description by means of an ontology. Then, all these profiles are then combined with the help of an ontology that establishes the possible relationships between all the elements present in a typical scenario of use in a digital library integrated in an e-learning environment. We have described the basic functionalities of the personalization system by means of use cases, and a methodology for building and ontology which describes the complete scenario of use. We have also identified the basic elements which are used to build such

ontology: user profiles, navigational profiles, user actions and the relationships established between these elements which are used by the personalization system.

Ontologies are a powerful tool for describing complex scenarios of use such as a digital library, where several concepts and relationships between these concepts can be identified and formally represented. The use of ontologies promotes the integration of new services into existing ones, and the interoperability with other systems through the appropriate semantic web services. New system functionalities and requirements can be added by including the appropriate description into the ontology framework that defines the digital library scenario of use.

Current and further research in this subject include the integration of the digital library personalization services with other personalization mechanisms provided by the virtual campus, towards a unique and complete user model. The digital library should become, therefore, another piece of the e-learning environment fully integrated into the learning process. The inclusion of new concepts related to the temporal validity of the ontology instances (resources, users and so) and their relationships should be also addressed. Finally, the definition of a validation rating algorithm combining both automatic but also user explicit rating systems is also under consideration.

References

Adams, K. (2002), "The Semantic Web: differentiating between taxonomies and ontologies", *Online*, Vol. 26 No. 4, pp. 20-3.

Adomavicius, G. and Tuzhilin, A. (1999), "Integrating user behavior and collaborative methods in recommender systems", *Proceedings of the CHI'99 Workshop "Interacting with recommender systems", Pittsburgh, PA*.

Arms, W.Y. (2002), *Digital Libraries*, The MIT Press, Cambridge, MA.

Brophy, P. (2001), *The Library in the Twenty-First Century*, Library Association Publishing, London.

Brophy, P. (2004), "Networked learning and networked information: towards a theoretical basis for the development of integrated information environment", *Proceedings of the 4th Networked Learning Conference, Lancaster, UK*, available at: www.shef.ac.uk/nlc2004/Proceedings/Symposia/Symposium9/Brophy.htm

Callan, J. and Smeaton, A. (2003), "Personalization and recommender systems in digital libraries", Joint NSF-EU DELOS Working Group Report, available at: www.dli2.nsf.gov/internationalprojects/working_group_reports/personalisation.html

Chen, H. (2003), "Towards building digital library as an institution of knowledge", NSF Post Digital Library Futures Workshop, Chatham, MA, available at: www.sis.pitt.edu/%7Edlwkshop/paper_chen.html

Colkin, E. (2001), "Personalization tools dig deeper", *Information Week*, No. 852, pp. 49-51.

Denny, M. (2002), "Ontology building: a survey of editing tools", xml.com, available at: www.xml.com/pub/a/2002/11/06/ontologies.html

Dix, A., Finlay, J., Abowd, G. and Beale, R. (1998), *Human-Computer Interaction*, 2nd ed., Prentice Hall, Englewood Cliffs, NJ.

Ford, N. and Ford, R. (1993), "Towards a cognitive theory of information accessing: an empirical study", *Information Processing & Management*, Vol. 29 No. 5, pp. 569-85.

Gapen, D.K. (1993), "The virtual library: knowledge, society and the librarian", in Saunders, L. (Ed.), *The Virtual Library Visions and Realities*, Meckler Corporation, Westport, CT, pp. 1-14.

Gruber, T.R. (1995), "Toward principles for the design of ontologies used for knowledge sharing", *International Journal of Human-Computer Studies*, Vol. 43, pp. 907-28.

Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J.T. (2004), "Evaluating collaborative filtering recommender systems", *ACM Transactions on Information Systems*, Vol. 22 No. 1, pp. 5-53.

Jonassen, D., Davidson, M., Collins, M., Campbell, J. and Haag, B. (1995), "Constructivism and computer-mediated communication in distance education", *The American Journal of Distance Education*, Vol. 9 No. 2, pp. 17-23.

Kasanoff, B. (2001), *Making it Personal*, Perseus Publishing, Cambridge, MA.

Lynch, C. (2003), "Digital library opportunities", *Journal of Academic Librarianship*, Vol. 29 No. 5, pp. 286-9.

Mor, E. and Minguillón, J. (2004), "E-learning Personalization based on Itineraries and long-term navigational behavior", *Proceedings of the 13th International World Wide Web Conference, Alternate Track Papers and Posters*, pp. 264-265.

Nielsen, J. (1998), "Personalization is over-rated", Alertbox. available at: www.useit.com/alertbox/981004.html

Riecken, D. (2000), "Personalized views of personalization", *Communications of the ACM*, Vol. 43 No. 8, pp. 27-8.

Rosenberg, M.J.E-L.e. (2001), *E-learning: Strategies for Delivering Knowledge in the Digital Age*, McGraw-Hill, New York, NY.

Sadeh, T. and Walker, J. (2003), "Library portals: toward the Semantic Web", *New Library World*, Vol. 104 Nos. 1184/1185, pp. 11-19.

Schonberg, E., Cofino, T., Hoch, R., Podlaseck, M. and Spraragen, S.L. (2000), "Measuring success", *Communications of the ACM*, Vol. 43 No. 8, pp. 53-7.

Sheth, A., Ramakrishnan, C. and Thomas, C. (2005), "Semantics for the Semantic Web: the implicit, the formal and the powerful", *International Journal on Semantic Web & Information Systems*, Vol. 1 No. 1, pp. 1-18.

Sicilia, M.A. and García, E. (2005), "On the convergence of formal ontologies and standardized e-learning", *Journal of Distance Education Technologies*, Vol. 3 No. 12, pp. 12-28.

Sinha, R. and Swearingen, K. (2001), "Comparing recommendations made by online systems and friends", *Proceedings of the Second DELOS Network of Excellence Workshop on Personalization and Recommender Systems in Digital Libraries*, available at: www.ercim.org/publication/ws-proceedings/DelNoe02/RashmiSinha.pdf

W3C (2004a), *eXtensible Markup Language (XML) 1.0 (Third Edition)*, available at: http://www.w3.org/TR/REC-xml/

W3C (2004b), *Description Framework (RDF) Primer*, available at: www.w3.org/TR/rdf-primer/

W3C (2004c), *Ontology Language (OWL) Guide*, available at: http://www.w3.org/TR/owl-guide/

Paper II

FERRAN, N.; CASADESÚS, J., KRAKOWSKA, M.; MINGUILLÓN, J. "Enriching e-learning metadata through digital library usage analysis", *The Electronic Library*, 2007, vol. 25, no. 2, p. 148-165.
ISSN 0264-0473

Afiliació dels autors:
Núria Ferran Ferrer
Information and Communication Sciences Studies, Universitat Oberta de Catalunya
Dr. Jaume Casadesús Brugués
Àrea de Tecnologia del Reg, Institut de Recerca i Tecnologia Agroalimentària (IRTA), Lleida
Dr. Monika Krakowska
Institute of Information and Library Science, Jagiellonian University, Cracòvia, Polònia
Dr. Julià Minguillón Alfonso
Computer Science, Multimedia and Telecommunications Studies,, Universitat Oberta de Catalunya

Resum en català:
A "Enriching e-learning metadata through digital library usage analysis" s'analitza les interaccions entre els perfils d'usuari I els recursos d'informació revelant dades interessants sobre l'ús dels recursos que poden ser afegides com a valors a metadades que descriguin aquests recursos, paliant l'allau d'informació i donant suport a l'avaluació de la qualitat dels recursos així com millorant el procés d'aprenentatge. Una proposta de camps de LOM (Learning Object Metadata) es proporciona i s'indica si es pot generar automàticament a partir de l'ús dels recursos que en fa una comunitat educativa.

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0264-0473.htm

Received 12 January 2007
Revised 18 January 2007
Accepted 19 January 2007

Enriching e-learning metadata through digital library usage analysis

Núria Ferran
Information and Communication Sciences Studies, Universitat Oberta de Catalunya, Barcelona, Spain

Jaume Casadesús
IRTA, Lleida, Spain

Monika Krakowska
Institute of Information and Library Science, Jagiellonian University, Krakow, Poland, and

Julià Minguillón
Department of Computer Science, Multimedia and Telecommunication, Universitat Oberta de Catalunya, Barcelona, Spain

Abstract

Purpose – The purpose of this research is to propose an evaluation framework for analyzing learning objects usage, with the aim of extracting useful information for improving the quality of the metadata used to describe the learning objects, but also for personalization purposes, including user models and adaptive itineraries.

Design/methodology/approach – The paper presents experimental results from the log usage analysis during one academic semester of two different subjects, 350 students. The experiment examines raw server log data generated from the interactions of the students with the classroom learning objects, in order to find relevant information that can be used to improve the metadata used for describing both the learning objects and the learning process.

Findings – Preliminary studies have been carried out in order to obtain an initial picture of the interactions between learners and the virtual campus, including both services and resources usage. These studies try to establish relationships between user profiles and their information and navigational behavior in the virtual campus, with the aim of promoting personalization and improving the understanding of what learning in virtual environments means.

Research limitations/implications – During the formal learning process, students use learning resources from the virtual classroom provided by the academic library, but they also seek information outside the virtual campus. All these usage data are not considered in the model proposed here. Further research is needed in order to obtain a complete view of the seeking information behavior of students for improving the users' profile and creating better personalized services.

Practical implications – This paper suggests how a selection of fields used in the LOM standard could be used for enriching the description of learning objects, automatically in some cases, from the learning objects usage performed by an academic community.

Originality/value – From libraries beginnings, they have been a "quiet storage place". With the development of digital libraries, they become a meeting place where explicit and implicit recommendations about information sources can be shared among users. Social and learning process interactions, therefore, can be considered another knowledge source.

Keywords Digital libraries, Learning, E-learning

Paper type Research paper

The Electronic Library
Vol. 25 No. 2, 2007
pp. 148-165
© Emerald Group Publishing Limited
0264-0473
DOI 10.1108/02640470710741296

Introduction

The intensive use of Information and Communication Technologies (ICT) such as the Internet increases the possibilities for both content searching and delivery. This new paradigm has completely changed the visions in the distance education field. For example, web based learning scenarios are becoming a common tool for both face to face and distance educational institutions. E-learning is one of the most promising and growing issues in the information society nowadays, mainly because of the growth of the Internet is bringing online education to people in corporations, institutes of higher education, the government and other sectors (Rosenberg, 2001). The growing need of continuous education and the inclusion of new multimedia technologies become crucial factors for this expansion. This fact is supported by two important issues: first, the apparition of new e-learning standards for describing complex learning scenarios, such as IMS-LD (IMS LD, 2003) and SCORM 2004 (ADL, 2004), and second, the new trends in education defined by what is known as the Bologna process (Bologna Declaration, 1999), where learners become the center of any educational experience, and all the activities, resources and scheduling are arranged according to each learner.

Nowadays, it becomes necessary to shift from heavily content-based courses to other formative actions where the activity is the key concept. Actually, activities and the competences developed by such activities are becoming the focus of any formative action. It is also important to promote the formal acknowledgement of skills, knowledge and competences gained through work experience, non-formal training and life experience, for prior learning recognition purposes. This setup promotes what is known as a lifelong learning scenario, where learners continuously improve their competences and knowledge by selecting the best formative activities according to their preferences, particularities and specific needs. Nevertheless, high quality educational contents become the basic pieces of this new learning process, but it is worth to remark that this learning process is user-centered, not content-centered.

In the last few years, personalization has become an important issue for both instructional designers and teachers. The high diversity of user profiles and backgrounds, and the new scenario defined by the Bologna Declaration makes necessary to personalize the learning process for each learner, according to his or her preferences, particularities, competences, and so. Content personalization allows teachers to describe adaptive itineraries depending on the necessities of each known profile, in order to deliver the appropriate content for each learner. Personalization is strongly related to user experience and satisfaction, which are supposed to be linked to academic performance and dropout rates, in the case of a virtual learning scenario. Furthermore, several studies have tried to discover the relevant attributes of the service quality in online environments, and in most of them, personalization was considered to be one of these essential attributes (Lee and Lin, 2005; Voss, 2003) among others such as quality of the content, usability or reliability.

In order to do so, providing adaptive itineraries needs from an underlying architecture where contents are highly structured and they can be properly stored, described and accessed in a dynamic framework. Learning object repositories (LOR) become a crucial element of any lifelong learning scenario, as they give support to the learning process. Therefore digital libraries, as providers of heterogeneous educational resources, are also essential. Actually, digital libraries are also a target for

personalization purposes, especially when they are integrated in a virtual learning environment (Ferran *et al.*, 2005).

In this paper we propose an evaluation framework for rating learning objects usage, with the aim of extracting useful information for improving the quality of the metadata used to describe such learning objects, but also for personalization purposes in adaptive systems (Towle and Halm, 2005). We also propose the use of an ontology for establishing the data collection process and ensuring a high degree of coherence between all the elements in the learning scenario, that is, user profiles, learning objects, interactions and so. Two study cases about the usage of several learning resources in a digital library related to a course in Statistics and a course in Archives Management are used to exemplify the relationships between user profiles, content delivery and personalization issues.

Learning objects usage in e-learning environments

In a lifelong learning scenario, learners could follow informal, non-formal and formal processes of learning (Coombs and Ahmed, 1974). In the first two options, the learning cycle can be based seeking through Internet and selecting the most suitable contents. Students in these processes do not need to get involved with any institution, but only for mentoring, evaluation and accreditation purposes. On the contrary, learners in a university receive, through its set of instructional designers, teachers, and librarians, pushed content with very few options to decide what, where and when study.

In both processes, there is the need for describing all the learning objects using the appropriate metadata. Learning online needs appropriate metadata for describing the content in order to retrieve and select it, and online learning needs metadata for describing not only the content in itself but also competences and even user preferences for building adaptive itineraries.

Learning objects can be considered as "any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning" (LOM, 2000). At an educational level, libraries have traditionally been the responsible institutions for organizing and providing metadata to these resources. If we focus on a digital library as a collection on educational material, it may be useful to think of it as a repository of learning objects. Basically, a repository stores educational resources and their descriptions for providing access and retrieval for teachers and students, either through Internet or with locked up access behind passwords within proprietary systems.

Librarians describe the resources of catalogues and other collections trough metadata in order to facilitate efficiently the delivery of information. The resource descriptions enable users to discover and identify existing materials and to evaluate and distinguish between different resources and allow the option to personalize information presented, via learner's information profile (Foster-Jones and Beazleigh, 2002). Metadata provide controlled and structured descriptions of resources through searchable access points such as title, author, date, location, description and subject, but can also provide interpretative information on the potential education application of resources or include described information about the relations between the resources (Friesen, 2002). The distinction of these two types of educational metadata is labeled as authoritative and non-authoritative (Recker and Wiley, 2001).

Unless authoritative metadata, which are generally managed by librarians, non-authoritative metadata are more likely to be generated by the final users of learning objects, so it is interesting to automatically produce these quality data by tagging the navigational actions performed by particular user profiles with learning objects, in order to meet students and teachers expectations when using a LOR. Regarding personalization purposes, LORs provide different services to users, from the searching/browsing possibilities to personal services such as a system that keeps track of user interests based on which educational resources he or she searches and downloads. For instance, SMETE is a learning object repository for teaching and learning of science, mathematics, engineering and technology at all levels, that includes a recommender system based on past user interactions (Neven and Duval, 2002).

Several techniques are used for guidance and for providing recommendations to users. Among others, collaborative filtering (Herlocker *et al.*, 2004) is one of the most successful ones. Briefly, collaborative filtering is selecting content based on the preferences of people with similar interests, basically by pooling and ranking informed opinions (or experiences of use) on any particular topic. That is to say, an automatic system collects information about user actions (explicit, such as voting or answering a question; or implicit, such as noticing which offered links are visited and which are not, how often and how much time is spend) and determines the relative importance of each content by weighting all the collected information among the large amount of users. Both navigational techniques are also valid in a digital library scenario. As stated in Fourie (2006), some authors found that personality types and learning styles will influence information-seeking styles (Limberg, 1999). Therefore, searching and browsing activities can be a useful source of information about user behavior.

Nevertheless, it seems clear that all the needed information cannot be stored into the learning objects in form of metadata, as they would become too specific, thus reducing reusability, the main goal of any learning object repository. The use of external structures for supporting these needs can be implemented by means of ontologies which describe the relationships between all the elements in the learning scenario. Ontologies can also be used to better describe such elements, incorporating a semantic level of information which can be used to enrich the learning process (Sicilia and García-Barriocanal, 2005).

The basic idea of this paper is that the interactions with learning objects carried out by different user profiles can be stored in a structured way, and then shared for future users with similar necessities overcoming information overload and difficult quality assessment. On the other hand, the analysis of such interactions may also reveal interesting facts about the use of learning resources that can be added as new metadata to such resources, improving their overall quality, but also in the learning process itself if it is described by means of formal descriptions, such as those provided by the IMS-LD standard, for example.

The UOC learning object repository

From a teacher's perspective, setting up a learning object repository and providing contents is not the most important issue in a learning scenario. On the contrary, the most interesting information is extracted from the usage that learners perform on such repository. The cost of setting up a repository of learning objects needs to be justified, at least, by a high degree of use, and by a continuous feedback that allows teachers and

instructional designers to extract useful information from the learning process followed by learners. Quality is also one of the major concerns for any repository, as learners will use it only if they feel confident of the available contents. Although quality is ensured by the institution setting up the repository, it is also important to allow users to participate in a continuous quality improvement process, both explicitly, by means of user ratings or annotations, or implicitly, by means of analyzing the usage of the repository and inferring relevant patterns.

In the particular case of the Universitat Oberta de Catalunya (UOC, in English known as Open University of Catalonia), a pure online university where students and teachers interact by means of a virtual campus, a shift from a content based towards an activity based learning process is now under development, although the pedagogical model was already designed under a user-centered approach. The UOC e-learning environment can be considered a lifelong learning scenario, where both contents and the learning process are provided by the institution, which ensures a high degree of quality control in a top-down approach.

There is evidence that teaching methods are shifting from a transmission of knowledge to a problem-based learning process. This increases the use of libraries, collections and repositories (Limberg, 1999). Following this quote, the UOC pedagogical model is based in a new model for teaching and learning where the teacher becomes a guide in the learning process, for which the student is ultimately responsible (Sangrà, 2003). For each course, the teacher establishes a learning plan where a calendar, an activity schedule, the basic communication tools and a suggestion of learning objects available at the institutional repositories are proposed.

These learning objects are stored in the digital library framework in two different repositories, depending on its source type. First, the "OPAC" repository (Online Public Access Catalog), where the recommended bibliography is stored, which is linked to the service that provides digital versions of chapters of available books. The OPAC also includes the subject textbooks in HTML and/or PDF formats. Second, the "Digital Collection" repository, with content from external providers subscribed by the academic library such as academic databases, electronic journals, as well as free Internet resources, proposed exercises and previous exams, or thesis and dissertations done by teachers and students of the university from previous semesters. According to Fox and Shalini (2002), the UOC learning object repository is a client-server based approach, in opposite to peer-to-peer approaches, as a basic policy for ensuring quality issues in the learning process. Nevertheless, all external resources used by learners are unknown, so it is important to understand that only a partial knowledge about the information behavior is available.

The available contents are located at the virtual campus, either in the digital library as a whole, or as a subset placed in each virtual classroom. Students and teachers have free remote access to the digital library, but only the students registered to one particular course have access to its specific library classroom. In any case, it is also well know that students use external information sources for accomplishing their learning goals, and teachers might also recommend the use of Internet search engines for doing so. In order to know the real implications of this fact, in a survey performed by the digital library on students enrolled on the first semester of the 2004-2005 course, the 31 percent of the students stated that they start seeking for information for

educational subject purposes from the library resources, while 53 percent start straight ahead from an Internet portal or a search engine (UOC, 2005).

The learning objects stored at the library are catalogued in MARC 21 if they are accessible trough the OPAC, or in Dublin Core if they are accessible from the Digital Collection. Currently now, there is an ongoing project at the university for cataloguing all the subject textbooks using the LOM standard, for satisfying the growing needs of the university, while the MPEG-7 standard is also under evaluation for description purposes (Pascual *et al.*, 2006).

Access and navigational profiles

Preliminary studies have been carried out in order to obtain an initial picture of the interactions between learners and the virtual campus, including both services and resources usage. These studies try to establish relationships between user profiles and their information and navigational behavior in the virtual campus, with the aim of promoting personalization and improving the understanding of what learning in virtual environments means.

It is remarkable to take into account the particularities of UOC students. The most common profile is an adult with an average age between 26 and 35 years old (57 percent), married and with children (55 percent), with a full time job (93 percent) and, a very important issue, who has got already a previous university degree (60 percent), but wants to be updated and improve his or her knowledge, either for personal or professional reasons. A study about the satisfaction of graduated students shows that they chose the UOC for the learning model as the main reason, because it is a fully distance online system that allows them to study from anywhere at anytime, and it is very flexible. The concept of lifelong learning is also important, as 38 percent of graduated students have also chosen the university because they wanted to improve their knowledge, and 44 percent of graduated students have chosen a degree related with their job because 28 percent of students wanted to improve in the exercise of their professions (UOC, 2005).

A first internal study (UOC, 2005) was carried out to determine the way learners interact with all the educational resources available through the virtual classroom, the digital library and Internet. A total of 1,108 students, covering 20 different degrees, were asked to participate into a survey where they behavior with respect to learning resources was analyzed. User profiling by means of segmentation was carried using the following variables: how often and from where (and how) they access to basic learning resources (R, including the subject textbook), the communication spaces in the virtual classroom (V), additional bibliography (B) and optional further resources and readings (F); how often they use these resources for preparing the continuous evaluation activities, and how often they use them for preparing the final validation test or exam. For example, there is a question related to place how and from where they access to the resources, which combines the use of computers (D, in form of HTML or PDF digital resources) or paper textbooks (T), from home (H), office (O) or public spaces (P, for example libraries or cybercafes), or while commuting (C). Each possible combination of values is dichotomized in order to show whether a combination is present or not, converting categorical scales into binary variables, grouping contiguous values. A total of 11 binary variables were relevant for clustering

purposes, using a non hierarchical typological analysis. User satisfaction with respect to the learning resources usage was evaluated according to this setup.

Four typologies were discovered, as shown in Table I. Capital letters are used to show a strong use or relationship, weak otherwise.

In the light of these results, it is clear that textbooks based in paper are still a very important piece in the learning process, probably, as some authors have stated, because development of e-books has been led primarily by technology instead of by users' requirements, and the gap between functionality and usability is sufficiently wide to justify the lack of success of the first generation of e-books. Therefore, it is crucial to identify needs and requirements of the target community so that the design can fulfill their needs and expectations. The acquisition of a well-defined user profile is an essential component of the design process for the successful development of e-books (Landoni and Diaz, 2003).

From Table I, we can see that half of the students (54 percent) use only the subject textbooks and read the messages posted in the classroom and that one third of students (36 percent) search and use additional educational resources such as bibliography and further readings. In order to prepare the assignments, all the students (95.5 percent) use the paper textbook which is sent to the students' home at the beginning of the semester. And only 8 percent of the students use always the bibliography accessible from the library classroom for preparing the assignments and only 10 percent the recommended by teachers. On the other hand, half of the students say that frequently (30.4 percent) or even always (19 percent) search into Internet for finding documents that will help them on preparing assignments.

Other additional variables used in this study showed also interesting information that can be used for personalization purposes. For example, students in the "Involved" typology where the oldest (around 42,5 years old), showing that age group might be used as a relevant variable for selecting resource types, as a reasonable indicator of previous study habits.

Another interesting study (Carbó *et al.*, 2005) related to the navigational behavior was carried out with students from several subjects from the Computer Science degree. In this study, the main goal was establishing a relationship between navigational behavior and academic performance, according to each scheduling (which is different for each subject). Preliminary results show that there are three different navigational patterns: first, students that connect every day or almost every day; second, students that mostly connect on weekends (from Friday to Sunday, both included); and third, students that only connect when they have to deliver an exercise or according to the published scheduling (i.e. to participate in a discussion in the virtual classroom forum). Navigational patterns and interaction level is strongly related with academic results, and very simple rules can be extracted from the interactions during the first week of a scheduled exercise, for example. These experiments were designed taking into account

Table I.
Hierarchical typological analysis results

Typology	Where	Format	Resources	Satisfaction
Standard (54%)	H, O	T	R, V	Very high
Explorers (25%)	H	D, T	R, V, F	Very high
Involved (11%)	H, o	D, T	R, V, B, F	High
Non-involved (9%)	h	T	R	Medium/low

user interaction mainly, but learning resources usage could be also another interesting parameter to include in such experiments, especially when learners are supposed to use specific resources provided by the teacher for solving an exercise.

Therefore, interactions must be studied at different levels, among all the different elements in the e-learning scenario, depending on the context being evaluated. When the use of learning resources is involved, it is important to gather information about which resources are accessed, when and, if possible, how are they evaluated by learners. This information could be used to improve resource visibility or ranking, according to previous experiences by similar users, for example.

Usage data harvesting and analysis

Our proposal consists in identifying and describing the elements in the learning scenario, namely users, resources and the learning process in itself, and establishing the relationships that occur between them during the basic interactions performed by users, extracting relevant information for our purposes.

In the case of the learning scenario of the Open University of Catalonia, students log in into the virtual campus and have access to several services. Among them, the mailbox service and the virtual classroom are the most important. In the virtual classroom (one for each subject a student is enrolled into), students find a teaching plan, a calendar, a set of learning resources and several notice boards and forums. Students are expected to follow the teaching plan according to the calendar, which guides them through all the learning activities they must perform, interacting with teachers and other students through the notice boards and forums, and using the selected learning resources but also other additional ones available through the digital library or external search engines. In an ideal scenario, the teaching plan is a dynamic learning process integrated in an intelligent tutoring system giving support to the virtual campus, providing students with adaptive learning itineraries (Mor and Minguillón, 2004).

Therefore, there are several types of interactions that might be relevant for analysis purposes in such scenario. The elements of the e-learning environment are learners, teachers, learning resources available through a repository, services (i.e. the digital library), and the learning process in itself, which can be seen as a complex dynamic service but as a special kind of resource as well. In this paper we are not interested in modeling interactions between students or between teachers and students, but those dealing with learning resources.

Following the ideas presented in Ferran *et al.* (2005), we propose to use an ontology for describing the learning process, as the core of the intelligent tutoring system. This ontology will use other sub-ontologies for handling all the interactions with the learning object repository, the interactions with other services and the user profiles. Each ontology is responsible to determine which information is relevant for usage analysis, and this information is shared among the different ontologies, making the learning management system (i.e. the virtual camps) aware of the interactions. For example, when students search for learning resources, the ontology responsible for the searching process in the repository uses information from the student profile, in order to select the most appropriate resources according to learning style and accessibility issues.

Building a user model

In order to build a multidimensional user model and feed it from usage data, several fundamental questions must be addressed, as stated in Smeaton and Callan (2005):

- What data should (and can) be collected and how can be captured?
- How are anomalous data recognized and filtered out?
- How should the data be analyzed and which parameters need to be set?
- How are data weighted appropriately over time?

The first question is partially answered through what is known as deep log analysis techniques (Nicholas *et al.*, 2006). Basically, it consists on triangulating and enriching data from all possible sources, namely campus navigational logs, library usage logs, socio-demographic data and academic background. These data is captured in different ways. Both navigational and library usage data is stored in web servers as log files, usually following a standard such as the Apache Common Log Format (CLF). Socio-demographic data and academic background is provided by students during the enrollment process, and it is updated each semester. All this information should be stored using a standard format (IMS LIP, 2005), in order to promote sharing with other institutions and services.

The second question involves the preprocessing stage of the collected data. Several common problems must be addressed: first, log server files are huge, around 50GB each week, with millions of lines to be processed, although less than 1 percent of the lines contain useful information for navigational pattern analysis. Even on a daily or hourly basis, performing such an analysis may be computationally prohibitive. This can be partially solved introducing specific marks in the web site and then filtering out those lines not containing such marks. This approach has two important advantages: first, resulting log files are much smaller, and second, as marks are directly related to user actions, it is much easier to track user real intentions. Other problems related to the use of log server files are the possible collisions for the IP addresses identifying each connection, because of the proxies used by the Internet providers. Nevertheless this is not a real issue if users are uniquely identified when they log in into the web site, as is the case of the UOC virtual campus, where each user session is uniquely identified and, therefore, it can be tracked for analysis purposes. Once preprocessing is done, all available data must undergo a data mining process using the appropriate tools, using proprietary but also ad-hoc software, in order to mine the raw data more sophisticatedly. Clickstream techniques (Mobasher *et al.*, 2002) try to discover navigational patterns that can be related to user tasks, by combining the identified actions in the preprocessed log files. On the other hand, simple statistical analysis can be also carried out to extract useful information, for example, which are the most common keywords used for locating a learning resource.

Finally, a particular challenge for personalization is that long-term models must encompass a timespan that is defined in terms of a human lifetime (Gemmel *et al.*, 2003; Smeaton and Callan, 2005). Furthermore, such models need also to incorporate the learning scenario scheduling, that is, the concept of academic semester in case of higher education, for example. This is important because user actions are determined by such scheduling. On the other hand, users in complex environments such as the UOC virtual campus receive multiple inputs from different spaces and services, showing different

behaviors depending on many events and variables (scheduling, experience of use, and so). Therefore, it is important to understand the real motivations that are the underlying cause for explaining user behavior. This can be partially accomplished by means of surveys and user tests, where quantitative but especially qualitative data about system usage is much better obtained. Nevertheless, this point is out from the scope of this paper, so it will not be developed here.

Experimental results

With the available data from one academic semester (from February to June 2006), an experiment has been performed in order to determine the usage of several learning resources available in the virtual classroom library space. We have chosen the user behavior data generated during the first assignment, where a few learning resources are supposed to be used by learners in order to solve the proposed learning activities. Our purpose was to determine the actions that the learners perform with the proposed learning resources in order to find information that will help in the learning design (LD) of next semesters as well as for improving the description of the learning objects, both of them objectives for personalization purposes.

We have analyzed the data from the usage performed by students with the learning objects available in two different virtual classrooms from two subjects, Statistics from the Computer Science degree, and Managing Archives from the Information Science degree. The Statistics subject had 280 students, while Managing Archives had only 60 students. Both subjects are mandatory for students if they want to get their degree. Learners have access to the resources through the virtual campus, while they are inside the virtual classroom. We do not analyze the specific usage of individual users, as we are interested in detecting typical behaviors. We use the log server files generated by the Apache web servers which act as front-ends for the virtual campus.

We will show a few examples of interesting student behaviors that we have detected that could help us to enrich metadata descriptions, even automatically, and we will also discuss the limitations of this method.

The statistics subject. In this case the students have to solve several exercises about descriptive statistics. They have available several examples similar to those exercises, and they also have a document with all the errata in the textbook, which is supposed to be read before the exercises are solved. Students also have a guide for planning the work to do which helps them to establish an appropriate pace for learning all the concepts needed to solve the exercises. They have almost three weeks for reading the guide, the textbook (incorporating the changes described in the erratum file), and using similar exercises as learning examples before they try to solve the exercises.

The usage data captured during these three weeks shows that some of the available examples are never used. One possible reason is that those examples are poorly tagged and then become "invisible" to students. Another reason is that those examples are not relevant for solving the first learning activity. In any case, this information could be useful for narrowing the searches performed by students when they try to find examples for this first learning activity. A "relevance" factor could be added to each learning resource with respect to each learning activity, according to the gathered data.

On the other hand, the erratum file is downloaded by most users, as expected, altogether with the learning guide. Therefore, the intelligent tutoring system could use this fact to warn students which have not done so (and probably their teachers and

tutors too) that the date for delivering the solution of the learning activity is coming in. This rule could be incorporated in the LD description of the learning activity and triggered by the ontology responsible of the learning process.

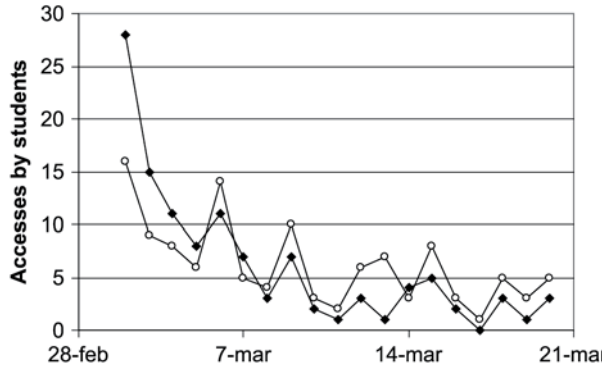
It is also interesting to remark that the erratum file has a different attitude with respect to the other resources: it is downloaded by the students following an exponential decrease; it is downloaded mostly at the beginning of the semester, whereas the rest of resources, for instance the learning guide but specially the additional exercises and examples, are accessed following teachers' recommendations or according to the teaching plan, as shown in Figure 1. Therefore, any usage analysis must be contextualized to the period of time where is carried on, because different results can be obtained.

The managing archives subject. On the other hand, in the subject "Managing Archives" we analyzed the behavior of students with another type of learning object, concretely exercises. Students had available in the virtual campus the statements in a file and the corresponding solutions in another.

The usage data confirmed that students downloaded the files in the order expected from the teaching plan. Indeed, in the first place students downloaded the statements, and after about two days, they downloaded the solutions (Figure 2).

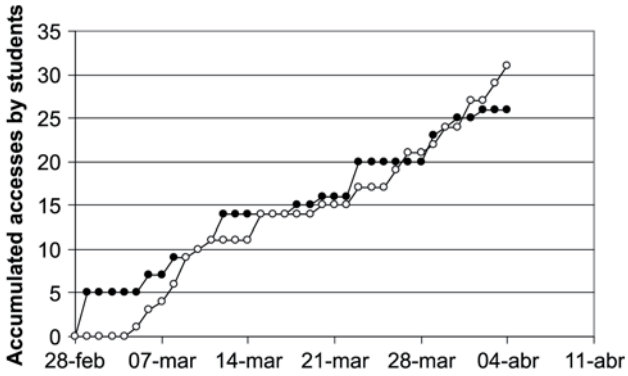
This type of test confirms that students effectively follow the teaching plan and that they try to solve themselves the problems before checking the solutions. A departure from this pattern would alert from negative studying habits. Also, each exercise statement can be rated by parameters such as its usage and the delay for accessing its solution.

Another possible utility of monitoring student access to resources is to find and rate LO. For instance, while a student is navigating in the class he or she may access the library or to Internet in order to find complementary resources. Here, the resources accessed by students of a given subject inside the virtual library of the UOC were effectively detected in our experiments. Therefore, this can help teachers discover other materials that could be included in the virtual library space of the classroom or added to the teaching plan.



Note: Closed symbols are the "Erratum file" and open symbols are the "Learning guide"

Figure 1.
Dynamics of students' accesses to different types of files



Note: Closed symbols are the "Statements" and open symbols are the "Solutions"

Figure 2.
Dynamics of accumulated students' accesses to the exercise files

These were some simple examples of students behavior when interacting with learning objects that appear on the virtual classrooms. What is important to note is that these observations were derived from raw data logged by default by web servers – i.e. no special utilities were required in the virtual campus for recording these raw data-. This means that the same approach can be used in other learning environments as well. However, this approach works at an HTTP session level and it is not able to identify and track the activities of a given student performed on different sessions. The anonymity of students implies that their behavior in a session can neither be related to previous sessions nor to other data from that student, such as his/her academic success.

In order to get a deeper view of the information student behavior, more detailed data is required and hence specific software intended for this purpose should be installed on the virtual campus. Further studies need to be carried on in order to have the complete view of the information behavior for e-learning purposes.

Enriching metadata in e-learning environments

Nowadays metadata used for describing the resources located in the virtual classrooms follow the Dublin Core initiative and are provided by librarians. When students retrieve these resources, they see title, author and other authoritative fields only.

Other metadata is used internally for librarians and lecturers for searching among the digital e-learning repository. Currently now, the repository cannot be completely accessible and searchable to the whole student community due to copyright laws that establish some restrictions for some resources, and only students enrolled into a particular subject are allowed to access such resources. Therefore, resources provided by the university in the virtual classroom are usually on a directory format rather, so metadata is not used by students for discovering the materials, although authoritative metadata (Recker and Wiley, 2001) can be automatically produced. Following the standard for Learning Object Metadata base schema (LOM, 2002) we suggest how a selection of LOM fields are involved in a process of metadata enrichment with information extracted from usage:

(1) LOM 1.5. *Keyword*: keywords or phrases describing the topic of the learning object. Through the searches performed by librarians and, when the repository is accessible for teachers and students, also by their searches, it is possible to rank keywords according to the number of times are used for retrieving the given learning object and, therefore, be used for improving recommendations or detecting misplaced keywords, for example.

(2) LOM 1.7. *Structure* and 1.8 *Aggregation level*: underlying organizational structure of the learning object. Usually, learning objects are described as independent chinks of information so they are considered to have an atomic structure, supposed to be indivisible. Usage data may reveal relationships with others learning objects and this fact can be used to create collections, hierarchical, linear or networked structures between them.

(3) LOM 3. *Metadata*: this field describes the metadata record itself. It could be used for registering all the automatic changes that the system performs, in order to further analyze the metadata enrichment process itself.

(4) LOM 5. *Educational*: this category describes the key educational or pedagogic characteristics of the learning object. Currently now it is one of the most criticized aspects of learning objects and LOM, as it is clearly underused. It is related with the quality of learning experience, so it becomes critical for any intelligent tutoring system dealing with learning objects for building adaptive itineraries:

- LOM 5.1. *Interactivity type* and 5.3. *Interactivity level*: these fields can indicate active learning, expositive or mixed, and its degree. These fields could be linked with the seeking information characteristic registered in the user profile. We could see, for example, that some resources provided in the "Statistic" course (additional exercises and examples) where not used by all users, as some of them preferred to base their study on the use of the hypertext material, which is more theoretical and textual instead of practical. Other users prefer learning-by-doing instead and, therefore, under a personalized learning process they could be recommended to use learning objects with such interactivity type and level.
- LOM 5.8. *Difficulty*: how hard it is to work with or through this learning object for the typical target audience. In an explicit way, users (the teacher but also the students) could suggest values for this field and the system could use it for selecting exercises to students according to their profile.
- LOM 5.9. *Typical learning time*: approximate or typical time it takes to work with or through this learning object for the typical intended target audience. For pure online learning objects (i.e. exercises with applets or simulations), the system can estimate the average time and use it to detect "outliers", that is, people that just walk through or people that spend too much time, taking the appropriate actions in each case.

(5) LOM 7. *Relation*: this category defines the relationship between this learning object and other learning objects, if any. This category can be used by the intelligent tutoring system to establish relationships between learning objects

according to their usage, especially those detected from the adaptive paths followed by students.

(6) LOM 9. *Classification*: this category describes where the learning object falls within a particular classification system. As an extension of the concept of keyword, it is possible to analyze the usage of the terms in the taxonomy for discovering interesting relationships and unused terms.

Regarding the learning process itself, which is handled by the intelligent tutoring system and the associated ontology, it is interesting to establish appropriate relationships between the proposed adaptive paths, their degree of acceptance by learners and their academic performance, as a valuable feedback for teachers and instructional designers. Following the same approach, these relationships could be incorporated in the metadata describing such itineraries following the IMS-LD standard.

Balancing privacy issues and social effects

Assurance and trust are considered the most important drivers of *e-service* satisfaction and loyalty (Reicheld and Schefter, 2000). Assurance results from perceptions of security, safety and trust. Security is the extent to which customers perceive the provider's web services to be free from intrusions by third parties, whereas privacy refers to the active maintenance of a level of confidentiality with respect to private information provided to the provider. Users expect they can trust organizations to protect any personal information they may have gathered. Preserving privacy and anonymity is an aspect that has been very important for libraries, and for that reason, an extended best practice of Integrated Library Systems is to destroy patron-related data. For example, the library only keeps the link between a patron and a book while the book is out on loan, in order to protect the library's holdings. But, once a book is returned to the library, all the patron activity history is deleted. This can explain why recommender systems are not likely to be used in libraries (Lynch, 2001).

On the contrary, a very important aspect that cannot be ignored is the fact that, in an Internet learning environment, users are always under control, in the sense that all actions are monitored and registered. This might seem a very invasive setup which harms user privacy and, therefore, undesirable. Nevertheless, there are several remarkable facts that need to be clarified:

- users know in advance that, in a web based environment, all actions are logged;
- the recommendation system must be designed in a non-intrusive manner and be user-friendly, including the possibility of disconnecting it or minimizing its participation in the browsing or searching activities; and
- the participation of individual users in the final recommendation system is completely anonymous.

Finally, it is also important to remark that the collected information is not meant for commercial purposes, and that the library (a non-profit organization) will use the data rationally and in a transparent way.

Furthermore, e-learning environments and digital libraries "can serve as meeting places where people can communicate with each other through the documents, annotations, and logs they make available to each other, and through the conversation

and discussion around this shared information" (Smeaton and Callan, 2005). This information exchange can be made in an implicit or explicit manner after the user consent in offering his or her usage data to the rest of the community. As usual, a tradeoff between personalization and privacy must be achieved (Kasanoff, 2001) in order to ensure the desirable social effects and a win-to-win scenario.

Conclusions

It seems clear that learning object repositories will become a basic piece of any learning environment, providing users with high quality contents, properly described and supported by means of metadata, taxonomies and ontologies. The integration of such repositories into the learning process is a key issue for ensuring a proper use, not just being a mere space where to find educational resources. E-learning success is somehow determined by the satisfaction in the learning process achieved by each student, and this satisfaction is directly related to the degree of interaction with learning resources, with the teacher and the other students in the virtual classroom, and the flexibility of the learning process in itself.

Furthermore, the forthcoming implantation of the Bologna process gives more responsibility to learners, making them the center of any formative action, promoting personalization, in order to adapt the learning process to each user particularities, needs and preferences, shifting towards a lifelong learning scenario. With the description of the learning process using competences and activities instead of contents, repositories need to be rethought in order to incorporate this new paradigm. Lifelong learning scenarios are based on a heavy use of available learning resources, where learners decide which contents are relevant for their purposes and which are not, with the possible guidance of an intelligent tutoring system. Discovering successful paths is a key issue for teachers and instructional designers for creating and updating such educational actions.

Content usage analysis is, therefore, a very important tool for ensuring all the learning resources in a learning process are properly used, satisfying the quality policies established by the institution, and providing system designers with the relevant information about the real use of the e-learning environment. We have described several experiments we have carried out with real usage data from an academic semester in two different subjects from two official degrees offered at the university, showing that even simple experiments reveal information about student behavior which can be incorporated into their learning profile but also into the metadata used for describing the learning resources and the learning process itself, in order to improve all the information handled by the intelligent tutoring system, the core of the personalization process. A proposal for enriching the LOM standard has been also described, showing the non-authoritative metadata fields that can be automatically generated from the analysis of the interactions between learners and resources.

Currently now, the UOC virtual campus is undergoing a major technological change for incorporating the new e-learning standards, such as IMS-LD, for example, with the aim of providing with formal descriptions the learning process, helping teachers and students to achieve their goals through a personalized learning process. Capturing external searches using Google or any other search engine, but also external database providers is also necessary to obtain a comprehensive view of the information seeking

behavior of students. In order to validate the proposal presented in this paper, we are setting up a repository for all the available resources in one of the subjects (concretely Statistics), which will be supported by an ontology which will capture all the actions performed by learners and will use this information to update the metadata used to describe such resources. We are also developing the ontology giving support to the intelligent tutoring system which will provide students with adaptive learning paths, according to their learning style and the interactions with the available resources.

References

ADL (2004), *Sharable Content Object Reference Model (SCORM) 2004*, 3rd ed., available at: www.adlnet.gov/downloads/300.cfm (accessed 8 November 2006).

Bologna Declaration (1999), *The European Higher Education Area. Joint Declaration of the European Ministers of Education*, convened in Bologna, 19 June.

Carbó, J.M., Mor, E. and Minguillón, J. (2005), "User navigational behavior in e-learning virtual environments", *Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 243-9.

Coombs, P.H. and Ahmed, M. (1974), *Attacking Rural Poverty: How Non-formal Education Can Help*, Johns Hopkins University Press, Baltimore, MD.

Ferran, N., Mor, E. and Minguillón, J. (2005), "Towards personalization in digital libraries through ontologies", *Library Management Journal*, Vol. 25 Nos 4/5, pp. 206-17.

Foster-Jones, J. and Beazleigh, H. (2002), "Metadata in the changing learning environment: developing skills to achieve the blue skies", *Association for Learning Technology Journal*, Vol. 10 No. 1, pp. 52-9.

Fourie, I. (2006), "Learning from web information seeking studies: some suggestions for LIS practioners", *The Electronic Library*, Vol. 24 No. 1, pp. 20-37.

Fox, E.A. and Shalini, U.R. (2002), "Digital libraries", *Annual Review of Information Science and Technology*, Vol. 36 No. 1, pp. 502-89.

Friesen, N. (2002), "Semantic interoperability, communities of practice and the Cancore learning object metadata profile", *Alternative Paper Tracks, Proceedings of the 11th World Wide Web Conference (WWW 2002), Hawaii*.

Gemmel, J., Lueder, R. and Bell, G. (2003), "The MyLifeBits lifetime store (demo description)", *ACM Proceedings of the 10th International Conference on the World Wide Web, Hong Kong*, pp. 1-7, available at: http://research.microsoft.com/~jgemmell/pubs/ETP2003.pdf (accessed 1 November 2006).

Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J.T. (2004), "Evaluating collaborative filtering recommender systems", *ACM Transactions on Information Systems*, Vol. 22 No. 1, pp. 5-53.

IMS LD (2003), *IMS Learning Design Specification*, available at: www.imsglobal.org/learningdesign/ (accessed 15 November 2006).

IMS LIP (2005), *IMS Learner Information Package Summary of Changes*, available at: www.imsglobal.org/profiles/lipv1p0p1/imslip_sumcv1p0p1.html (accessed 1 July 2006).

Kasanoff, B. (2001), *Making It Personal: How to Profit from Personalization without Invading Privacy*, Perseus Books Group, Cambridge.

Landoni, M. and Diaz, P. (2003), "E-education: design and evaluation for teaching and learning", *Journal of Digital Information*, Vol. 39 No. 4, available at: http://jodi.ecs.soton.ac.uk/Articles/v03/i04/editorial/ (accessed 7 November 2006).

Lee, G.G. and Lin, H. (2005), "Customer perceptions of e-service quality in online shopping", *International Journal of Retail & Distribution Management*, Vol. 33 No. 2, pp. 161-76.

Limberg, L. (1999), "Experiencing information seeking and learning: a study of the interaction between two phenomena", *Information Research*, Vol. 5 No. 1., available at: http://informationr.net/ir/5-1/paper68.html (accessed 1 July 2006).

LOM (2000), *LOM (Learning Object Metadata) working draft v4.1*, available at: http://ltsc.ieee.org/doc/wg12/LOMv4.1htm (accessed 1 July 2006).

LOM (2002), *Draft Standard for Learning Object Metadata*, available at: http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf (accessed 7 November 2006).

Lynch, C.A. (2001), "Personalization and recommender systems in the larger context: new directions and research questions", keynote paper presented at 2nd DELOS Network of Excellence Workshop on Personalisation and Recommender Systems in Digital Libraries, Dublin, Ireland, June 18-20, available at: www.ercim.org/publication/ws-proceedings/DelNoe02/CliffordLynchAbstract.pdf (accessed 1 July 2006).

Mobasher, B., Dai, H., Luo, T. and Nakagawa, M. (2002), "Discovery and evaluation of aggregate usage profiles for Web personalization", *Data Mining and Knowledge Discovery*, Vol. 6 No. 1, pp. 61-82.

Mor, E. and Minguillón, J. (2004), "E-learning personalization based on itineraries and long-term navigational behavior", *Proceedings of the 13th International World Wide Web Conference, New York, NY, May 19-21, 2004*, pp. 264-5.

Neven, F. and Duval, E. (2002), "Reusable learning objects: a survey of LOM-based repositories", *Proceedings of the 10th ACM international conference on Multimedia*, pp. 291-4, available at: www.cs.kuleuven.ac.be/cwis/research/hmdb/publications/files/Lorsurvey.pdf (accessed 1 July 2006).

Nicholas, D., Huntington, P. and Watkinson, A. (2006), "Scholarly journal usage: the results of deep log analysis", *Journal of Documentation*, Vol. 61 No. 2, pp. 248-80.

Pascual, M., Ferran, N. and Minguillón, J. (2006), "Integration of multimedia content and e-learning resources in a digital library", *Proceedings of the SPIE, Internet Imaging VII*, Vol. 6061, pp. g1-g11.

Recker, M.M. and Wiley, D.A. (2001), "A non-authoritative educational metadata ontology for filtering and recommending learning objects", *Journal of Interactive Learning Environments*, Swets and Zeitlinger, Lisse.

Reicheld, F.F. and Schefter, P. (2000), "E-loyalty: your secret weapon on the web", *Harvard Business Review*, Vol. 78 No. 4, pp. 105-13.

Rosenberg, M.J. (2001), *E-learning: Strategies for Delivering Knowledge in the Digital Age*, McGraw Hill, New York, NY.

Sangrà, A. (2003), "Universitat Oberta de Catalunya. A newly created institution", in D'Antoni, S. (Ed.), *The Virtual University. Models and Messages. Lessons from Case Studies*, International Institute for Educational Planning, Unesco, Paris, available at: www.unesco.org/iiep/virtualuniversity/files/uoc.pdf (accessed 1 July 2006).

Sicilia, M.A. and García-Barriocanal, E. (2005), "On the convergence of formal ontologies and standardized e-learning", *Journal of Distance Education Technologies*, Vol. 3 No. 2, pp. 12-28.

Smeaton, A.F. and Callan, J. (2005), "Personalisation and recommender systems in digital libraries", *International Journal on Digital Libraries*, Vol. 5 No. 4, pp. 299-308.

Towle, B. and Halm, M. (2005), "Design adaptive learning environments with learning design", in Koper, R. and Tattersall, C. (Eds), *Learning Design. A Handbook on Modeling and Delivering Networked Education and Training*, Springer, Dordrecht, pp. 215-26.

UOC (2005), *User Profile Surveys*, internal technical report, Universitat Oberta de Catalunya, Barcelona.

Voss, C. (2003), "Rethinking paradigms of service: Service in a virtual environment", *International Journal of Operations & Production Management*, Vol. 23 No. 1, pp. 88-104.

Further reading

Velterop, J. (2004), "The myth of 'unsustainable' open access journals", *Nature*, available at www.nature.com/nature/focus/accessdebate/10.html (accessed July 2006).

About the authors

Núria Ferran Ferrer is a lecturer at the Open University of Catalonia (UOC) and associated lecturer at the Universitat Autònoma de Barcelona (UAB) at the department of Information Sciences and Communication. Courses conducted: Digital libraries, Information sources applied to several areas, Information systems. Núria has degrees in Journalism at UAB (1998), in Documentation at UOC (2003), a Master in Information and Knowledge Society from IN3-UOC (2005) and is currently a PhD student at UOC in information behavior studies in e-learning students. Núria is the corresponding author and can be contacted at: nferranf@uoc.edu

Jaume Casadesús is a researcher on sensor systems for data acquisition at IRTA. He has a PhD in Biology from Universitat de Barcelona (1995), Bachelor in Computer Systems Engineering from UOC (2004), Master in Software Engineering from Universitat Politecnica de Catalunya (2006).

Monika Krakowska has a PhD and is an assistant at Jagiellonian University in Krakow, Poland in the Institute of Information and Library Science. Research interests include Information and Communication Technology, Information Literacy, Multimedia, New tools in Information Science. She is the Socrates Programme coordinator in the Institute of ILS JU. Her PhD degree is in Bibliology, Information and Library Science (2006).

Julià Minguillón received his PhD degree from the Universidad Autònoma de Barcelona (UAB) in September 2002, with thesis work about machine learning by cascading limited depth decision trees. In January 2001 he joined the Universitat Oberta de Catalunya (UOC) where he is a faculty member. Since November 2006, he is the Vice Director of the Internet Interdisciplinary Institute (IN3). He has developed learning resources for object oriented programming, abstract data types engineering and compiler construction. He is also involved in the integration of e-learning standards into virtual e-learning environments, such as LOM, SCORM and IMS-LD. His main research interests include the formal description of the learning process by means of ontologies, personalizing the learning process by means of adaptive itineraries based on reusable learning objects, and user modeling in virtual e-learning environments applying web mining techniques. He leads the PERSONAL project, the framework that articulates all the aforementioned research lines. He is also in charge of the UOC participation in the OLCOS (Open Learning Content Observatory Services) EU funded project.

To purchase reprints of this article please e-mail: **reprints@emeraldinsight.com**
Or visit our web site for further details: **www.emeraldinsight.com/reprints**

Paper III

FERRAN, N.; GUERRERO-ROLDÁN, A.; MOR, E.; MINGUILLÓN, J. User Centered Design of a Learning Object Repository", M. Kurosu (Ed.): *Human Centered Design, HCII 2009, Lecture Notes in Computer Science*, no. 5619, p. 679–688. ISBN 978-3-642-02805-2

Afiliació dels autors:

Núria Ferran Ferrer

Information and Communication Sciences Studies, Universitat Oberta de Catalunya

Ana-Elena Guerrero-Roldán

Computer Science, Multimedia and Telecommunications Studies, Universitat Oberta de Catalunya

Dr. Enric Mor

Computer Science, Multimedia and Telecommunications Studies, Universitat Oberta de Catalunya

Dr. Julià Minguillón Alfonso

Computer Science, Multimedia and Telecommunications Studies, Universitat Oberta de Catalunya

Resum en català:

"User Centered Design of a Learning Object Repository" té per objectiu el disseny d'un repositori d'objectes d'aprenentatge centrat en l'usuari. Un repositori ha de perseguir el desenvolupament i l'adquisició de competències genèriques i específiques, com les competènices informacionals. En aquest article s'ofereix els resultats d'una primera fase del procés de disseny centrat en l'usuari i proporcionen informació per conèixer i entendre requeriments dels usuaris i de les tasques així com el context d'ús dels usuaris. L'article presenta els resultats de dos estudis, un de qualitatiu i un de quantitatiu. A partir d'aquests estudis es pretén dibuixar el comportament informacional que inclou l'accés, tractament, integració, avaluació, creació i comunicació de la informació per finalitats educatives que permetrà aportar dades per integrar, tenint en compte als usuaris i els seus comportaments, els repositoris d'objectes d'aprenentatge als entorns virtuals d'aprenentage.

User Centered Design of a Learning Object Repository

Nuria Ferran[1], Ana-Elena Guerrero-Roldán[2], Enric Mor[2], and Julià Minguillón[2]

[1] Information and Communication Science Studies
[2] Computer Science, Multimedia and Telecommunications Studies
Universitat Oberta de Catalunya, Rambla Poble Nou 156, Barcelona, Spain
{nferranf,aguerreror,emor,jminguillona}@uoc.edu

Abstract. This work outlines the design process of a user centered learning object repository. A repository should foster the development and acquisition of both generic and specific informational competencies. The results of the first stage of the user centered design process are presented which provide a clear understanding of user and task requirements and the context of use. A user study was conducted using quantitative and qualitative methodologies. A qualitative approach was performed through the content analysis of 24 in-depth interviews achieved through a random stratified sampling method. Regarding the quantitative approach, more than 5 million student navigation sessions were processed in order to know the real information behavior accomplished in the virtual campus and more specifically all the services and resources used and the search actions carried out by users. Our aim is to achieve a thorough informational behavior analysis that involves access, treatment, integration, evaluation, creation and communication of information for learning purposes which will be useful for integrating learning object repositories in virtual learning environments.

Keywords: Learning Object Repository (LOR), user centered design, log analysis, content analysis, e-learning, information-related competencies.

1 Introduction

Learning object repositories are becoming a common tool for organizing educational content, that is, all the resources used in a learning process. The main aims of a repository are, firstly, to assure access to content and its conservation and, secondly, to promote a high degree of reusability of the available resources. But learning is much more than just contents, and there is a real need to integrate the use of learning object repositories (LOR) as part of the learning process so that students take advantage of the new possibilities it offers. Using the learning object repository should be in itself a true learning experience.

Furthermore, the new European Higher Education Area (EHEA) promotes the design of learner-centered processes that focus on the acquisition and development of competences rather than on the consumption of contents. One of the goals of the EHEA framework is to 'create' professionals with appropriate skills that help them to manage in the current information society. Thus, informational-related competencies become crucial. Learning object repositories are therefore one of the most important

M. Kurosu (Ed.): Human Centered Design, HCII 2009, LNCS 5619, pp. 679–688, 2009.
© Springer-Verlag Berlin Heidelberg 2009

elements of any e-learning system and the repository interface must be designed keeping in mind that real usage must be captured in order to provide instructors with data that describes the interaction between learners and the repository, thus the abovementioned learning experience can be measured to some extent. The integration of the learning object repository as part of the virtual learning environment will not be complete unless a user-centered design approach is taken on board. This paper is organized as follows: Section 2 describes learning object repositories and their integration in e-learning systems. Section 3 describes the definition of a user-centered design project to build a learning object repository. Section 4 presents the first stage of such a project; the research of the user in the context of a virtual learning environment. Finally, the conclusions drawn from this work and future research lines are summarized in Section 5.

2 Learning Object Repositories

With the creation of the new European Higher Education Area (EHEA) [1] and for a better alignment of learning with the requirements of the knowledge society, the education sector needs a new model of learning [2]. In order to do so, teachers must change their role from dispensers of knowledge to facilitators of individual and collaborative learning and knowledge development. This means a transition from an educational model based on established information channels to a new model where there are diverse channels [3]. Textbooks, workbooks, lectures and other pre-digested information from lectures must bring about a learning process based on information resources available in the real world [4]. Students should be encouraged to select resources from the Net, use/reuse and share them with the rest of the academic community.

A Learning Object Repository (LOR) is a basic service that provides learners with the contents they need according to the learning context in any moment of the learning process [5]. Besides, it helps teachers and instructors to better manage all the available resources and to understand the real usage of these resources by learners. Furthermore, there is a huge generation of selected or created resources in each semester, therefore the LOR becomes the right tool for managing all these contents.

In addition, the academic community is acting as a curator of the quality of the chosen/created learning objects, as teachers and students are putting into practice the appropriate information competencies. In this sense, learning object repositories are perceived as an essential tool for such a collaborative learning approach. On the other hand, learners must have some basic skills for accessing the repository and the interaction with it should increase their informational skills. The basic operations related to a learning object repository (from the learner's point of view) are information searching, browsing and retrieval. Users should be empowered to have the right information, at the right time, in the right format, with the optimal quality to meet a specific information need that fulfils a learning goal.

As stated in [6], the learning repository is in itself not enough to ensure a successful learning experience. It is necessary to build a true learning community around the learning object repository, with the aim of maintaining a continuous process of creating, sharing and reusing educational resources. In a formal learning scenario, such as

that defined by the new EHEA paradigm, the learning object repository cannot be just another technological service provided by the virtual learning environment, as it will be probably ignored by most of its potential users, mainly learners. Preliminary experiments [7] show that learners tend to minimize their interaction with static resources as they can be downloaded once and used locally many times. Therefore, it is important to integrate the learning object repository as an active element of the learning environment, promoting its use among learners. In order to do so, we propose to adopt a user-centered design approach, analyzing the real context where the learning object repository has to be integrated and taking into account the real user requirements from a methodological point of view.

3 User Centered Design of a Learning Object Repository

This work is part of a large project that takes place in a higher education institution, the Open University of Catalonia (UOC), with the aim of promoting the development and acquisition of competencies through the use of learning object repositories. The UOC is a purely online university that is currently evolving towards the EHEA. It has more than 40,000 students and more than 2,500 staff including instructional designers, teachers, tutors, academic and technical staff. The UOC uses a virtual campus as an integrated e-learning environment that allows students to pursue their studies purely online. We intend to design and develop a LOR that is not only useful as a mere repository but, at the same time, its use becomes an active element of the learning process, so students using the repository will achieve a set of competences.

The university established a set of organizational requirements mainly related with its technological architecture and also related with the e-learning methodological model. The second set of requirements is user defined and will be obtained through the user research described later on.

The current technological infrastructure is based on DSpace, which is already in use at the institution for publishing research results. This is the typical use of such technology, although DSpace [8] is intended to be used for storing and managing learning resources, that is, to be the core of the institutional learning object repository. One of the main drawbacks of DSpace is its user interface, which needs to be completely redesigned in order to be really usable. The default user interface reproduces the internal structure based on communities, sub-communities and collections, where each item in a collection is identified by its author, its title or its keywords. Although this may be sufficient for most typical uses, nevertheless, in the case of educational resources it is not so clear. For example, exercises do not have a clear title. On the other hand, some resources are created collaboratively during the academic semester, so the figure of the author is not clear. Therefore, we need to redefine the key elements that will be used for browsing the learning object repository, according to the desired learning goals.

As our intention is that the learning object repository is a true learning experience, its main goal cannot be just providing learners with searching and browsing capabilities. Quite the opposite, learners are expected to develop informational competences while they use the learning object repository, together with the acquisition and development of competences related to the repository thematic subject (if any, i.e. Statistics). In order to do so, we will redefine in full the user interface in order to provide a

comprehensive browsing experience, which will help learners to establish a relationship between resources, topics, keywords, competences and so. In this sense, it is worth remarking the MACE project [9], which aims to provide new user interfaces for browsing. We intend to extend the capabilities of the MACE project search engine in order to accommodate the new requirements imposed by the EHEA paradigm. Furthermore, we also intend that the virtual learning environment gathers and analyzes real usage data, in order to provide learners with not only a more improved personalized system but also teachers and managers with a better understanding of the learning process built around a learning object repository.

3.1 UCD to e-Learning

Taking into account all these elements, we decided to plan and develop this project applying user-centered design (UCD) [10]. The UCD is both a design philosophy and a product development process. This discipline places the user at the center of all the process, taking into account their characteristics, needs and wishes. As a mode of philosophy, the UCD is based on the principle that a key element for the success of a product is its adaptation to the user. Adaptation is understood at different levels which include the adaptation to human characteristics and limitations, adaptation to users' needs and desires, adaptation to the context of use. A UCD product development process includes three main phases: gathering user requirements, designing the product iteratively and finally, evaluating the prototypes of each design iteration.

By applying a UCD process to the design of the LOR, we can ensure that the repository provides what it was initially conceived to provide and, at the same time, we hope to obtain important results on the Learner-Centered Design which is how the design should be done in order to guarantee a good learning experience. The project follows the principles of the ISO 13407 [11], namely the active involvement of users and a clear understanding of the user and task requirements, an appropriate allocation of functions between users and technology and the iteration of design solutions and multi-disciplinary design. This international standard describes four user-centered design activities:

1. Specification of the context of use: identify the users of the LOR and under what conditions they will use it.
2. Requirement specification: identify students' needs and goals and organizational requirements.
3. Creation and development of design solutions: these designs will take into account the information gathered in the two previous phases.
4. Design evaluation: designs are evaluated taking into account users, requirements and the context of use.

The following section presents the results of the first and second phases.

4 User Research and Specifications of the Context of Use

Keeping in mind the user context and the application needed, we have used a multiple methodology approach on the information behaviour. On the one hand, a qualitative

perspective through content analysis and discourse analytical methods, and on the other, a quantitative approach through log analysis and data mining. By combining both methodologies we get, the reasons why students develop an information behaviour profile and their real navigational behaviour which provides data without the bias as their actions were transparently recorded.

The virtual campus is an integrated e-learning environment which includes all the needed services and tools. The learning object repository is one of its elements. As the main users of the learning object repository, learners are the subject of the user research.

4.1 User Qualitative Research

The virtual campus becomes a common space where students develop information behaviour; that is to say, how they execute a set of activities, such as the identification of needs, the search for, use and transfer of information [12]. And it is in this learning environment where students acquire or are able to identify, find, evaluate, organise, communicate and use the information effectively, both for solving problems and for lifelong learning [13]. Therefore, in the research carried out, we approach students in order to describe their information behaviour and their information-related competencies in the academic context in order to improve the services used to access and use the information needed for achieving the learning goals.

In this study, 24 in-depth interviews were performed on mature e-learning students from a purely virtual university (UOC). Throughout course 2006-07, there were a total of 38,842 students enrolled in undergraduate programmes, where the average age was between 26 and 35 years (58%) and more than 68% of the total students had a full time job and 55% had children.

Keeping this typology of student in mind, a stratified random selection of a sample was made with segmentation according to categories on a series of variables. We distinguished three age groups: a group of 25-35 year olds which includes the student average age; a group of 35-45 year olds and finally a group of more than 45 year olds as this age is regarded as critical for the digital divide [14]. The population was segmented in terms of gender (female/male) which can be considered proportional as there are 51% male to 49% female. Afterwards, we segmented each stratification in two groups that we called “Novice” and “Advanced” students. “Novice” referred to those students that have only 1 or 2 semesters at the university; therefore they have some knowledge and skills of the virtual campus resources. On the other hand, we used “Advanced” for those who have been enrolled for 3 or more semesters which are supposed to be more information competent. Finally, we stratified the segmentations once again in two subgroups “Experts” and “Non-experts” in terms of Information-related competencies. “Experts” was used for students that we considered to be information competent and therefore fulfilled two conditions: the first one is that they had at least once searched and retrieved an electronic article from any of the subscribed databases of the digital library and secondly, they had at least once uploaded content on the Internet, i.e. videos, photos or created a weblog.

This stratification was not proportionate, in the sense that the number of samples in each category did not necessarily correspond to their relative size in the population.

This was not regarded as a problem once the goals of the stratification are justified by the objectives of the research [15].

The initial stratified sampling was designed through telephone interviews in order to hold the face-to-face meetings. Prior to the meeting, they received a small questionnaire by email about when was the last time they searched and used information for their several needs in their academic, workplace and daily-life environment. This step helped participants to get a previous idea of what the interview was about and provide the interviewers with an incident case to be explored during the interview in order to make participants remember past real information-seeking situations [16].

Finally, the structured interviews were performed; each of them lasting from 60 to 90 minutes, between September and October 2007, in Barcelona, Spain. These interviews were recorded by audio and video, with the permission of the participants. Afterwards, the interviews were transcribed in text form and used as the raw data for the content analysis method. Afterwards, all this data was human-codified with a software tool, NVIVO 7.0 [17]. We created a codebook with all variable measures which were established following the main actions that the information behaviour manifested: access, treatment, integration, evaluation, creation and communication [18].

4.2 User Quantitative Research

The goal of the quantitative research is to collect data about how the students use the e-learning system in everyday real situations. This usage data will be processed and analyzed to obtain new evidence about system usage and about student navigation. To do that, a three-level methodology of analysis has been used [19] focusing at the first level. Working at the first level, obtaining navigation paths, requires a complex system for managing and processing log files. To obtain patterns and other results typical data mining methods are used [20].

The data set used in the analysis belongs to the spring semester of course 2006-07. Throughout the semester, the log files have been gathered, filtered and have been stored to be able to be processed in order to obtain the student navigation paths. The first step of the analysis is to obtain all the navigation sessions of the students. Afterwards, these navigation paths or sessions will be analyzed with the goal to discover new information about student behavior and system usage.

The log files do not provide information rich enough to obtain relevant information about the student behavior in the system. Therefore, to be able to obtain the navigation behaviour of the users, we decided to introduce a set of embedded marks in the system [19]. These marks leave a clear track in the log files and can be processed later. To do that, a marking strategy has been designed obtaining a map of marks that have been embedded in the e-learning system.

The spring academic semester had a duration of 136 days, the first day being the 28th February and lasted until the dates of the publication of the final marks being the 12th and 13th of July. All students share the same virtual learning environment and also the key dates of the course. The number of students registered during this semester is 29,531. During the course, the log files from the active front-end servers of the virtual campus were received and pre-processed everyday. The log files, once pre-processed thus eliminating all that redundant, incomplete or superfluous information, occupy a disk space of more than 150 GB. From this point, these files have been

processed through an algorithm that allowed obtaining all the navigation sessions. These sessions have been stored in a unique file of a size of around 790 MB.

All the student navigational sessions obtained from processing the data set have been stored in one file where each line represents a single user navigation session in the virtual campus. This file contains 5,326,697 lines and, therefore, the total number of sessions of the semester. Studying carefully the data, it can be observed that there are some incomplete. We decided to eliminate them since they do not bring information about user navigation and probably were failed sessions. Once eliminated, the data set obtained contains 5,293,237 sessions which is more than 99% of the original file.

These navigation sessions are a very valuable information source that had never been obtained before. They show a lot of information about the users and how they use the virtual campus. Even so, it is interesting to continue processing and analyzing this data. As a matter of fact, these navigation paths represent the starting point of new analyses and studies about the virtual learning environment and its users.

5 Results

The analysis of the interview content and the log analysis showed two main student behavioral patterns. As a result of the qualitative analysis, one major interpretative repertoire of information behaviour appeared among mature e-learning students following the “googlizated behaviour” usually applied to teenagers [21]. Contrary to that, these students are not using the teenager’s tools but paper as a preferred format, Word software as a creation tool and e-mail as their communication tool.

From the log file analysis, we can draw what we called a "blackberrized behavior” in the sense that that students access to services and tools of the campus with the aim of getting updates. The common session of a student last 7 seconds and there are several sessions each day. This finding reinforces the idea that students do not use the virtual campus for studying but for communication purposes.

Some of the recurring expressions describing both patterns found have been organized from the codification performed through the following information actions:

a) Accessing to information:

Google is the search tool for all the interviewers: "when I have to search anything "I do" a Google, nowadays, it is the main source of information." The common opinion is that Google is the most complete source of information and that it retrieves information in a very fast way and that it is well structured and thematically ordered.

Related to accessing the virtual learning environment, the student behavior is very constant in the sense that we can conclude that during working days, the duration of the navigation session is short and the access is frequently. 7 seconds is the time one user needs to login, load the virtual campus’ home page, and make a glance to see if there are new messages. This value is important to be taken into account for interface design. If we want to capture the student’s attention in the home page, the system should be capable of generating and show useful information in less than 7 seconds. The distribution of the duration of the sessions shows that there is another operation a part from see the news and to go out, and it is the one where students carry out tasks in the virtual campus and, therefore, with a longer duration.

b) Treatment of information:
Generally, people for study purposes print their selected information sources or didactical materials. Paper was considered more secure, provided a feeling of more control, something more familiar, and something more convenient. But this feature was different among the ages of interviewers and their discipline. Navigation data showed that only a small group of students use the digital library as a study habit, reinforcing the fact of not studying online and on screen.

c) Evaluation:
In the academic context, students are provided with the necessary content to carry out the evaluation tasks. If they want more information, it is something complementary, but within the virtual classroom, they got the material provided by the most reliable source which for them was the teacher. Adding to that, the analysis of the virtual campus navigation sessions showed that there is no clear pattern. Students can be grouped in several navigation profiles because they behave differently depending on the moment of the day and on the moment of the course. Therefore, educational interfaces have to take into account behavior changes throughout the semester and also the information needs of each course.

d) Creation and Integration:
The main tool for creating and integrating information is Word software. Students mostly when generating a document for evaluation purposes recognize that "I copy and paste material that I found on the Internet and then I create the essay in this process on a Word file". So this way of cutting and pasting is the most universal and familiar method of study rather than reading and digesting. Furthermore, these text files are also used for storing bibliographic sources.

e) Communication:
In order to interact with other fellows and with teachers, email is the main channel. However, they like collaboration among students from their class but they hate the "forum" or "debate" tool provided in the classroom and they also dislike teamwork as they associate it to having to meet in a face-to-face manner. As mentioned, most of the students log in on the virtual campus several times each day only to check their mail.

6 Conclusions

This present study presents the work in progress of the integration of a learning object repository in a virtual learning environment from a user- centered design approach. Learning object repositories are becoming a key element in virtual learning environments as they provide the basic infrastructure for managing all the learning resources used during the learning process. Nevertheless, it is important to ensure that learners will use the learning object repository as expected, and that such interaction will be captured in order to be further analyzed.

Once the first stage of the user-centered design process is finished and the requirements are analyzed, it is then time to proceed with generating design solutions. In the particular case of a virtual learning environment and taking into account the user research, the learning object repository must be useful not only as a simple space

where learning resources are found, but as an active component providing learners with a true learning experience. Students are "googlized" and "blackberrized" and, therefore, the interface must allow them to use the repository taking into account their actual skills but should be formative enough to change their behavior and improve their skills. Obviously, this learning experience will be different depending on the nature of each subject, but some common requirements can be identified. We want the learning object repository to help learners to establish relationships between resources, in accordance with their similarities, overlapping and even user preferences. It is in this process of establishing relationships that the learner creates a mental map of the whole subject, thus improving his or her understanding of it. The browsing engine should avoid the use of Google-like search boxes and promote other interactive elements such as tag clouds and hierarchical taxonomies, among others.

Current and future research in this subject should include the development of a social layer with regard to the learning object repository, as part of its deeper integration within the learning process, in order to promote its continuous use, analyzing user behaviour in order to detect possible problems or improvements. On the other hand, as the number of learners accessing the virtual learning environment through mobile devices is increasing, hence, it is necessary to adapt some of the services provided by the learning object repository to this new learning scenario, taking into account mobility and accessibility issues. This is not only a technological issue, because the access device has different purposes depending on the context where is used. Finally, the adoption of semantic web techniques will enable better personalized services combining all the elements in the learning process, that is, users, contents and services. Personalization is one of the key aspects in providing learners with a true learning experience and which is perceived as being something real and useful.

Acknowledgments. This work has been partially supported by Spanish government projects PERSONAL(ONTO) ref. TIN2006-15107-C02 and E-MATH++ ref. EA2008-0151.

References

1. The Bologna Declaration, http://www.bologna-bergen2005.no/Docs/00-Main_doc/990719BOLOGNA_DECLARATION.PDF
2. Geser, G. (ed.): Open Educational Practices and Resources - OLCOS Roadmap 2012. Open eLearning Content Observatory Services (2007)
3. Benito Morales, F.: Nuevas necesidades, nuevos problemas. Fundamentos de la alfabetización en información. In: Gómez Hernández, J.A. (coord.) Estrategias y modelos para enseñar a usar la información. Murcia, KR (2000)
4. Breivik, P.: Student learning in the information age. ACE. Oryx Press, Arizona (1998)
5. Ferran, N., Mor, E., Minguillón, J.: Towards personalization in digital libraries through ontologies. Library Management Journal 25(4/5), 206–217 (2005)
6. McNaught, C.: Are Learning Repositories Likely To Become Mainstream In Education? In: Proceedings of the 2nd International Conference on Web Information Systems and Technologies, Setubal, Portugal, April 11-13, pp. IS9–IS17 (2006) (Keynote address)

7. Ferran, N., Casadesús, J., Krakowska, M., Minguillón, J.: Enriching e-learning metadata through digital library usage analysis. The Electronic Library 25(2), 148–165 (2007)
8. DSpace, http://www.dspace.org/
9. Mace Project, http://portal.mace-project.eu/ProjectSearch
10. Norman, D.A., Draper, S.W.: User Centered System Design; New Perspectives on Human-Computer Interaction. Lawrence Erlbaum Associates, Inc., Mahwah (1986)
11. ISO/IEC 13407 Human centred design processes for interactive systems, http://www.usabilitynet.org/tools/13407stds.htm
12. Wilson, T.D.: Recent trends in user studies: action research and qualitative methods. Information Research 5, 3 (2000)
13. ASSL Information literacy standards for student learning, http://www.ala.org/ala/mgrps/divs/aasl/aaslproftools/informationpower/InformationLiteracyStandards_final.pdf
14. Katz, J.E., Rice, R.E.: Social consequences of Internet use. Access, involvement and interaction. The MIT Press, Cambridge (2002)
15. Neuendorf, K.A.: The content analysis guidebook. Sage Publications, Thousand Oaks (2002)
16. Talja, S., Keso, H., Pietiläinen, T.: The production of context in information seeking research: a metatheoretical view. Information Processing Management 35, 751–763 (1999)
17. Richards, T., Richards, L.: Using computers in qualitative research. In: Denzin, N.K., Lincoln, Y.S. (eds.) Collecting and Interpreting Qualitative Materials, pp. 211–245. Sage Publications, London (1998)
18. Léveillé, Y.: Les six étapes d'une démarche de recherche d'information. In: La recherche d'information à l'école secondaire, Ministère de l'Éducation du Québec, Direction des ressources didactiques (1997), http://pages.infinit.net/formanet/cs/chap2.html
19. Mor, E., Minguillón, J., Garreta-Domingo, M., Lewis, S.: A Three-Level Approach for Analyzing User Behavior in Ongoing Relationships. In: HCI Applications and Services. 12th International Conference, HCI International, Proceedings, Part IV, Beijing, China, July 22-27. LNCS. Springer-, Heidelberg (2007)
20. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern classification. Wiley, New York (2001)
21. Manuel, K.: Teaching information literacy to Generation Y. Journal of Library Administration 36(1/2), 195–217 (2002)

Paper IV

FERRAN, N.; PÉREZ-MONTORO, M. La Gestión de la Información Personal en usuarios intensivos de las TIC, *El Profesional de la Información*, 2009, vol. 18, no. 4, p. 365-373.
ISSN 1386-6710

Afiliació dels autors:
Núria Ferran Ferrer
Estudis de Ciències de la Informació i la Comunicació, Universitat Oberta de Catalunya
Dr. Mario Pérez-Montoro
Facultat de Biblioteconomia i Documentació, Universitat de Barcelona

Gestión de la información personal en usuarios avanzados en TIC

Por **Núria Ferran-Ferrer** y **Mario Pérez-Montoro**

Resumen: *Se presenta un modelo de análisis del comportamiento informacional global de un colectivo de individuos (estudiantes de la Universitat Oberta de Catalunya) que tienen una percepción positiva sobre el uso de las tecnologías de la información y la comunicación y que realizan un uso intensivo de las mismas. A partir de una aproximación cualitativa, mediante 24 entrevistas y un posterior análisis del contenido, se identifican cuatro perfiles distintos de gestión de la información personal (reactivo, pasivo, exhaustivo y proactivo) en base a diez variables subyacentes (acceso, gestión y usos de la información, competencias informacionales, perfil cognitivo, actitud, percepción de las TIC, ámbito académico, profesional y de la vida diaria) y se ponen de relieve las diferencias de comportamiento informacional dependiendo del ámbito en el que se encuentren. La identificación de los perfiles es un estadio básico del diseño centrado en los usuarios que facilita la realización de intervenciones específicas para cada tipo de usuario, respetando requerimientos de herramientas y procesos para que puedan desarrollar su comportamiento informacional de forma eficiente y eficaz.*

Palabras clave: *Comportamiento informacional, Modelos, E-learning, Búsqueda de información, Gestión de información, Uso de información, Contextos de uso.*

Title: Personal information management of ICT's intensive users

Abstract: *A large part of the research on personal information management is scattered through many journals due to the intrinsic characteristics of this topic. This work tries to compile a more complete, global and integrated study of that type of management strategies. Our study offers an analytical model of the informational behaviour of a group of individuals (students of an e-learning university, UOC) who have a positive perception of the information and communication technologies and use them intensively in their academic, professional and everyday contexts. From this qualitative research (24 interviews and content analysis), four user profiles (reactive, passive, comprehensive and proactive) have been identified in relation to personal information management and its relevant variables: access, management and use of information, cognitive style, attitude, information-related competencies, ICT perception, and academic, professional and everyday contexts. These profiles show the differences in information behaviour for each context (daily life, professional and academic). Identification of the user profiles was the essential first step in a user centered service design that addresses the specific intervention/contribution appropriate to each user type, taking into account tool and process requirements needed for them to develop efficient information behaviours.*

Keywords: *Information behaviour, E-learning students, Patterns, Usage contexts, Information seeking behaviour, Information management, Information use.*

Ferran-Ferrer, Núria; **Pérez-Montoro, Mario**. "Gestión de la información personal en usuarios avanzados en TIC". *El profesional de la información*, 2009, julio-agosto, v. 18, n. 4, pp. 365-373.

DOI: 10.3145/epi.2009.jul.02

Artículo recibido el 23-04-09

Aceptación definitiva: 16-06-09

***Núria Ferran-Ferrer**, licenciada en periodismo por la Univ. Autònoma de Barcelona, en documentación por la Univ. Oberta de Catalunya (UOC) y máster en sociedad de la información por IN3-UOC, es profesora titular en la UOC en los estudios de ciencias de la información y la comunicación en la Licenciatura y Grado de Documentación desde 2005. Ha participado en proyectos de investigación europea sobre tecnologías de la información en relación con la gestión de patrimonio (Coine) y con contenidos educativos abiertos (Olcos). Está realizando el doctorado sobre el comportamiento informacional de los estudiantes de una universidad virtual y la transferencia de competencias informacionales entre los contextos académico, profesional y de la vida diaria.*

***Mario Pérez-Montoro** es doctor en filosofía y ciencias de la educación por la Univ. de Barcelona (UB) y máster en organización de sistemas de documentación en la empresa por la Universitat Politècnica de Catalunya. Ha realizado estudios de postgrado en el Istituto di Discipline della Comunicazione de la Università di Bologna (Italia) y ha sido profesor visitante (visiting scholar) del CSLI (Center for the Study of Language and Information) de la Stanford University (California, EUA). Su trayectoria científica se centra sobre algunos aspectos (conceptuales, semánticos, epistemológicos y pragmáticos) relacionados con la teoría de la información y la gestión del conocimiento. Es profesor de la Facultat de Biblioteconomia i Documentació de la UB.*

1. Introducción

La expresión "gestión de la información personal" suele utilizarse en la literatura especializada con dos sentidos diferentes: como actividad y como disciplina de estudio (**Jones**, 2007). Como actividad se corresponde con todas las acciones que un individuo suele realizar para adquirir, crear, organizar, almacenar, recuperar, utilizar y distribuir la información necesaria para completar las diferentes tareas y las responsabilidades que tiene asumidas a nivel personal, social y laboral. Como disciplina, en cambio, se encarga del estudio de ese tipo de acciones que se corresponden con esa actividad concreta y se alimenta, entre otros, de presupuestos provenientes de otras disciplinas como la ciencia cognitiva, las ciencias de la información, la inteligencia artificial y la interacción persona-ordenador.

Como objeto de estudio, el tratamiento de la información personal se centra en dos frentes complementarios. Por un lado, investiga los diferentes tipos de comportamientos que suelen manifestar los individuos a la hora de organizar su información personal. Y, por otro, esta disciplina se encarga también de proponer estrategias y metodologías para que los individuos puedan acercarse, en la medida de lo posible, a la mejor de las situaciones (o situación ideal) respecto a la gestión de esa información personal: disponer de la información exacta, en el lugar adecuado, en el formato correcto, en el momento justo, y lo suficientemente completa y de calidad como para satisfacer todas las necesidades de información que surgen en los distintos ámbitos de la vida diaria.

La evolución de la investigación en el campo de la gestión de la información personal se está viendo frenada por un gran obstáculo que se deriva de la propia naturaleza del objeto de estudio abordado: la fragmentación (**Jones**, 2007; **Jones**; **Teevan**, 2007). Habitualmente la información personal que maneja un individuo a diario se encuentra fragmentada de formas diferentes. Por un lado, se presenta en diferentes formatos (en papel y en formato electrónico, principalmente). Por otro, se encuentra distribuida en diferentes tipos de documentos (como, por ejemplo, textos, fotografías, música o vídeos). Y, por último, esta diversidad de formatos y tipos de documentos ha generado la aparición de una serie de estrategias y herramientas parciales que han contribuido a la consolidación de la gestión fragmentada de esa información personal.

Esta situación ha provocado también cierto grado de disgregación en la propia investigación sobre este tema. Así, gracias a la revisión bibliográfica sobre la gestión de la información personal publicada en el *Annual review of information science*, conocemos investigaciones enfocadas exclusivamente en las aplicaciones utilizadas, como el e-mail. Otros abordan, de forma exclusiva, el uso de los favoritos y del historial en el navegador o la búsqueda y recuperación de información pública y en la Red. También hay estudios sobre la organización de documentos en papel y en formato electrónico. Otros se han centrado en el análisis de ciertas estrategias de almacenamiento (ver citas en **Jones**; **Teevan**, 2007).

Nuestro trabajo intenta evitar esta manera fragmentada de abordar la gestión de la información personal y ofrecer una visión más completa, global e integrada de ese tipo de estrategias de gestión. Concretamente, se presenta un análisis del comportamiento informacional global de un colectivo de individuos (estudiantes de la *Universitat Oberta de Catalunya, UOC*, una universidad virtual) que tienen una percepción positiva de las tecnologías de la información y la comunicación y que realizan un uso intensivo de las mismas. El estudio se ha realizado en los tres ámbitos principales de sus vidas diarias: el entorno académico, el profesional y el privado. El objetivo principal del trabajo es comprobar si, respecto a las estrategias de tratamiento de la información personal, hay perfiles diferenciados entre esos sujetos, cómo podrían caracterizarse los mismos y cuáles son los requerimientos de herramientas y procesos para que puedan llevar a cabo su comportamiento informacional.

2. Metodología

El estudio se centró en los usuarios (**Dervin**; **Nilan**, 1986) y se utilizaron métodos cualitativos. Primero se llevó a cabo una breve entrevista telefónica para conseguir la muestra, y a continuación se realizaron 24 entrevistas en profundidad y se contextualizaron mediante la técnica del estudio de incidencia (*incident case*[1]) (**Talja**, 1999). Posteriormente se realizó un análisis de contenido sobre las transcripciones textuales de las entrevistas.

2.1. Entrevista telefónica para la estratificación y el estudio de incidencia

En el curso 2007-08, la población de estudiantes matriculados en la *UOC* era de 40.264 alumnos. Esta universidad utiliza de forma intensiva las tecnologías de la información y la comunicación como vehículo de aprendizaje mediante su campus virtual. La media de edad de estos estudiantes estaba entre 26 y 35 años (53,8%), más del 85% tenían trabajo a tiempo completo y el 55% tenía un familiar a cargo.

Teniendo en cuenta estos tipos de estudiantes, se procedió a seleccionar una muestra estratificada de acuerdo con las variables de interés de la investigación. La estratificación no fue proporcional a la población, y el número de participantes fue de 24, porque la representatividad no se requería en la parte cualitativa del estudio (**Neuendorf**, 2002).

"En el entorno laboral, el perfil predominante es el disciplinado, y en el privado es el caótico"

Se establecieron tres grupos de edad para organizar la muestra mediante los datos proporcionados por la universidad y por las entrevistas telefónicas. Se tuvo en cuenta la media de edad citada y la edad de 45 años como crítica respecto al fenómeno de la brecha digital (**Katz**; **Rice**, 2002). Esta población se segmentó utilizando el criterio del género (51% masculino / 49% femenino); según su experiencia universitaria virtual (novatos: estudiantes que han cursado sólo de 1 a 2 semestres / avanzados: estudiantes que han cursado más de 2 semestres en la *UOC*) y también se contempló su nivel de competencias relacionadas con la información (expertos/no-expertos). Se consideró que un estudiante pertenecía a la categoría de expertos si se cumplían dos condiciones: la primera era si había buscado y recuperado algún artículo en las bases de datos de la biblioteca y la segunda condición era si había colgado contenido en internet (vídeos, fotos, etc.). Entre los 24 estudiantes se consiguió una representación del área de las ciencias sociales (derecho, comunicación, económicas, psicología), de las humanidades (historia, filología) y de la tecnología (informática y telecomunicaciones).

La entrevista telefónica se utilizó también para introducir la técnica del estudio de incidencia que serviría de ayuda posteriormente a la hora de realizar las entrevistas. Esta técnica consiste en identificar comportamientos informacionales que han ejemplificado los entrevistados en el pasado para luego utilizarlos a modo de ilustración y aclaración en el momento de la entrevista. Se les enviaron por correo electrónico tres fichas, una para cada contexto de búsqueda, tratamiento y uso de la información (profesional, académico, privado) que debían rellenar describiendo la última acción que habían realizado con información en esos contextos. De esta manera se obtuvieron estudios de incidencia establecidos previamente por los entrevistados en cada contexto de acción: académico, profesional y privado. Uno de los ejemplos del ámbito privado que se citaba en más fichas era cómo guardar las fotografías de las vacaciones de verano y cómo encontrar un lugar para publicarlas y compartirlas o cómo hacerlas llegar a otras personas. Estas fichas se utilizaron durante las entrevistas para contextualizar las preguntas en casos reales de búsqueda, tratamiento y uso de la información.

Edad (años)	Competencias informacionales	Femenino		Masculino	
		Novatos	Avanzados	Novatos	Avanzados
25-35	Experto	1	1	1	1
	No experto	1	1	1	1
35-45	Experto	1	1	1	1
	No experto	1	1	1	1
Más de 45	Experto	1	1	1	1
	No experto	1	1	1	1

Tabla 1. Estratificación de la muestra de los 24 participantes

"En el entorno académico, el perfil disciplinado y el caótico son equivalentes"

2.2. Entrevistas en profundidad y análisis del contenido

Las entrevistas se realizaron en el mes de octubre de 2007 en Barcelona. Duraban entre 60-90 minutos y se registró, previa autorización, tanto el sonido como el vídeo. Previamente a las 24 entrevistas finales, se realizó una prueba piloto.

El guión de las entrevistas era semiestructurado siguiendo las diferentes acciones que compone el comportamiento informacional para la gestión de la información personal: adquirir, crear, organizar, almacenar, recuperar, utilizar y distribuir la información necesaria para completar las diferentes tareas y las responsabilidades que tiene asumidas a nivel personal, social y laboral.

Finalmente se transcribieron las entrevistas en formato texto y se utilizaron estos datos para un análisis de contenido utilizando el software *Nvivo* de soporte a las metodologías cualitativas (**Richards**, 1999).

La transcripción textual de las entrevistas se utilizó como variable de muestreo para realizar la investigación. Se procedió a realizar el análisis de contenido identificando las unidades de registro a partir de las variables que recogemos en la tabla 2 (**Krippendorf**, 1990).

Mediante las variables 1, 2 y 3 se dibuja el comportamiento informacional global desde la necesidad de información para poder cumplir con los objetivos o tareas (**Léveillé**, 1997).

La cuarta variable se identificó porque, una vez realizadas las entrevistas presenciales, se hizo evidente que el atributo personal relacionado con las competencias informacionales obtenido con las entrevistas telefónicas no recogía completamente la totalidad del conocimiento, habilidades y actitudes que tenían los participantes en los tres escenarios para encontrar, eva-

Variables de estudio
1. **Acceso a la información** (necesidades / fuentes / formas de acceso – suscripciones, etc.)
2. **Gestión de la información** (tratamiento e integración / recuperación y preservación / criterios de selección / criterios de satisfacción)
3. **Usos de la información** (creación / integración / canales de comunicación / compartimiento de información / formas de trabajo colaborativas)
4. **Competencias informacionales** (experto / no-experto)
5. **Perfil cognitivo** (caótico / disciplinado)
6. **Actitud** (espectador / implicado)
7. **Percepción de las TIC** (entusiasta / realista / crítico)
8. **Ámbito académico**
9. **Ámbito profesional**
10. **Ámbito vida diaria**

Tabla 2. Cuadro de clasificación

luar y utilizar información de forma efectiva (**Virkus**, 2006).

También se analizaron otros aspectos relacionados con el comportamiento informacional como es la personalidad del sujeto (**Palmer**, 1991). Este aspecto se codificó dentro del perfil cognitivo: los disciplinados son sujetos metódicos, organizados y, en cambio, los caóticos son poco organizados e improvisan sus propias reglas para conseguir los resultados.

La sexta variable, la actitud, es uno de los componentes de las competencias informacionales. En el proceso de aprendizaje y en el proceso de búsqueda de información, el concepto actitud va ligado a la motivación. Esta variable permitió codificar los comportamientos de los identificados como implicados y como espectadores. Los primeros, a diferencia de los espectadores, están motivados por aprender, por las acciones y los procesos y no buscan una recompensa o evitar un castigo para empezar a actuar (**Dörney**, 1994).

La quinta variable hace referencia a la percepción de las TIC. Los entusiastas ven las fortalezas de la Red y consideran que les proporciona poder y libertad. Los realistas, en cambio, no dan a las TIC prioridad alguna, las ven como ventajosas únicamente en situaciones específicas. Y, finalmente, los críticos tienen sus reservas respecto a las TIC. Esta clasificación fue usada por **Savolainen** en su análisis sobre los usos de internet (2004).

El resto de las variables (8-10) caracterizan los contextos en los que tiene lugar el comportamiento informacional (**Wilson**, 1981).

La codificación fue realizada por dos técnicos que no habían intervenido en el proceso de las entrevistas, tal y como sugiere **Krippendorff** (1990). Y como paso final se realizó el test de fiabilidad para cada variable, que permitía comprobar el índice de acuerdo entre los codificadores. En este estudio la fiabilidad (0,83) está por encima de los índices recomendados por **Krippendorff** e incluso por encima de 0,80 (alpha), nivel que permite llegar a conclusiones sólidas y fundamentadas, más allá de las especulaciones.

3. Resultados

Los resultados que se presentan a continuación son los obtenidos de las variables determinantes para dibujar la gestión personal de la información de los sujetos estudiados. En este sentido, particularidades personales como el género, la edad, el número de semestres matriculados en la *UOC*, o las responsabilidades familiares no proporcionaron diferencias en relación con el tema.

Para presentar los resultados, seguiremos la estructura planteada en la tabla 2. Los valores de algunas variables se presentan en términos de frecuencias absolutas, es decir, el número de incidentes encontrados en la muestra. También se proporcionan tabulaciones cruzadas de varias dimensiones para mostrar las relaciones entre variables. Los valores de algunas variables se presentan de forma textual para favorecer la apreciación de la riqueza de matices.

3.1. Acceso a la información

En el entorno personal, aprovechando que las entrevistas se realizaron justo después de las vacaciones de verano, el acceso a la información se relacionó básicamente con la reserva de un viaje y todos los casos de incidencia se basaron en este tema. Las necesidades de información y las fuentes consultadas fueron bastante homogéneas entre los participantes. Las fuentes a las que más acudió la mayoría de los entrevistados fueron las webs obtenidas utilizando *Google*, portales temáticos, fuentes personales mediante el teléfono, *Wikipedia*, webs de comercio online, y prensa en papel y digital. Algunos participantes citaron bibliotecas públicas, *Emule*, *YouTube*, audiolibros y blogs. Respecto a cómo actuaban para conseguir estar al día sobre la información que les interesaba coincidieron de forma generalizada en que su estrategia no pasaba por subscribirse o sindicarse a las fuentes de internet. En sus propias palabras, eran ellos lo que "controlaban internet" y no era la Red la que los dominaba a ellos.

En el entorno profesional, en cambio, el acceso a la información proporcionó unos resultados más homogéneos. Las principales fuentes de información fueron personales (teléfono, encuentros presenciales o correo electrónico). Otras fuentes secundarias que se indicaron hacen referencia a la prensa y a bases de datos. Algunos entrevistados comentaron tener intranet y *newsletters*

en el trabajo. Y otros, en menor proporción, destacaron que a los trabajadores de su rango no les estaba permitido consultar internet pero habitualmente coincidían con que la mayoría de los aspectos relacionados con la información se realizaban mediante la Red.

El acceso a la información en el entorno académico se realizaba en todos los casos desde los materiales didácticos proporcionados por la universidad en cada aula virtual, material de foros y tablones del campus y el contenido prescrito por el profesorado. A nivel equiparable a este material predefinido por el plan docente, se citó a *Google*, y *Wikipedia* como fuentes de inicio de búsquedas. Se indicaron también fuentes especializadas como bases de datos (sobre todo los estudiantes de derecho) y visitas a bibliotecas presenciales. En algunos casos se enumeraron portales de estudiantes para compartir apuntes y exámenes, entre ellos el más citado fue el creado por los estudiantes de la *UOC* que contiene un repositorio de exámenes. Muchos entrevistados coincidieron con la afirmación de uno de ellos: "la *UOC* te da las herramientas y el camino marcado, no necesito buscar mucho fuera".

3.2. Gestión de la información

La valoración de las fuentes no fue una acción que se practicase habitualmente en el ámbito privado, no se considera muy necesaria. El sistema del navegador o sitios web especializados (como *Del.icio.us*) en gestionar los enlaces favoritos no fueron en ningún caso los instrumentos utilizados para este fin sino que, como alternativa, para llevar a cabo esa gestión crean una libreta o un fichero de texto.

Las facturas se guardaban en papel aunque con la banca electrónica muchos se han pasado al archivo digital. Las hojas de cálculo se utilizan para contabilidad y para realizar comparativas de precios de compras online. Los contenidos más habituales que podemos encontrar en sus archivos son las fotografías. No acostumbran a nombrar o titular cada una de esas fotografías, sólo las archivan por paquetes y cronológicamente en carpetas a las que sí asignan un título. Los contenidos bajados por internet no se renombran y en muy pocos casos se clasifican en profundidad; por lo general se archivan en carpetas a las que se les asigna un nombre o título según el producto cultural que contenga (música, películas) y, en otros casos, según el género. Las copias de seguridad o *backups* son una práctica extendida según comentaron los entrevistados.

Los criterios de selección de las fuentes fueron homogéneos en el contexto profesional. Todos coincidieron en que accedían sólo a fuentes muy conocidas y habituales. Manifestaban que no podían arriesgarse y existía mucha desconfianza a lo desconocido. Las aplicaciones utilizadas para tratar, integrar y guardar la información forman parte del paquete *Office* de *Microsoft*. El "corta/pega" de internet a *Word* era la técnica más habitual, así como la impresión en papel. Acciones relacionadas con la preservación no fueron destacadas puesto que en la mayoría de los lugares de trabajo estas responsabilidades no recaen en los trabajadores.

El prestigio del autor de una fuente, la fecha de actualización, el diseño de la página, reseñas sobre la fuente, entre otros, destacan entre los numerosos criterios comentados que se tienen en cuenta a la hora de valorar las fuentes de información. La publicidad de los sitios web se percibe como un elemento relacionado con la poca veracidad de las mismas.

En relación con las búsquedas internas, la impresión y el almacenamiento, el formato preferido por los estudiantes es el pdf frente al html. Aprovechando las herramientas informáticas, realizan esquemas y resúmenes. En algún caso aislado indican la utilización de un gestor bibliográfico. La estructura de los archivos con los que trabajaban es homogénea (desde el punto de vista temático y cronológico) ya que la propia universidad proporciona buena parte de los contenidos didácticos en formato CD para facilitar la preservación de los materiales de las asignaturas matriculadas. Por último, es interesante remarcar que suelen utilizar las hojas de cálculo como instrumento para crear bases de datos y para la gestión del tiempo de estudio y la entrega de los trabajos de evaluación continuada.

3.3. Usos de la información

En el contexto privado, el envío de fotografías para compartirlas con amigos y familiares es uno de los usos más habituales. En muy pocos casos se citaba un repositorio específico (*Picasa*, mayoritariamente) para compartir este tipo de información. El teléfono es el dispositivo preferido para comunicarse. *Skype* sólo se utiliza en casos de familiares lejanos. Para chatear, el *Messenger* es el sistema más usado. Y para comunicarse con el resto de los habitantes de la casa, la plataforma más común eran las notas en papel colgadas en la nevera. En la forma de compartir y difundir información relacionada con las aficiones personales es dónde se detectó un nivel más avanzado de competencias informacionales al utilizar herramientas de la Web 2.0. También se comentó el conocimiento de licencias como las *Creative Commons* y se explicitaban las plataformas utilizadas para la sindicación de contenidos. *Emule* y *YouTube* fueron las plataformas más populares identificadas.

Los intercambios interpersonales presenciales eran el canal más habitual para difundir la información en el entorno profesional. El e-mail está en segundo lugar, con *Outlook*, y normalmente se utiliza la opción de "acuse de recibo". Las intranets, los boletines electró-

nicos de comunicación interna y las reuniones presenciales, también se citaron. *Messenger* se usaba en las empresas entre compañeros cuando no había un chat en la intranet. Del paquete *Office* se utilizan intensivamente *Word*, *Excel*, *Access* y *PowerPoint*.

Cabe destacar que las entrevistas se realizaron en octubre de 2007 y en aquel momento aplicaciones o servicios como *Gmail* y *Facebook*, entre otros, no estaban tan extendidos como en la actualidad.

3.4. Competencias informacionales

Mediante la entrevista telefónica los participantes fueron catalogados en 50% expertos y 50% no-expertos respecto a las competencias informacionales. Pero con la codificación del contenido a partir de contextos de uso destacaron 11 personas susceptibles de ser consideradas competentes en el entorno de trabajo y en el privado simultáneamente. En el ámbito académico sólo pudieron considerarse expertas 5 personas.

Contexto	Competencias informacionales	
	Expertos	No-expertos
Privado	15	9
Profesional	16	8
Académico	5	19

Tabla 4. Resultados de la variable competencias informacionales

3.5. Actitud

No todos los participantes tenían la misma actitud en los tres contextos analizados. Los que tenían una actitud diferente podían identificarse como implicados en los entornos profesionales y privados; en cambio, en el académico fueron considerados como espectadores.

Contexto	Actitud	
	Implicados	Espectadores
Privado	12	12
Profesional	14	10
Académico	12	12

Tabla 5. Resultados de la variable actitud

Contexto	Perfil cognitivo	
	Disciplinados	Caóticos
Privado	10	14
Profesional	20	4
Académico	12	12

Tabla 6. Resultados de la variable perfil cognitivo

3.6. Perfil cognitivo

Los resultados obtenidos muestran que el comportamiento informacional era muy diferente según el contexto. En el entorno laboral, el perfil predominante era el del disciplinado, mientras que en el privado era el caótico el más abundante. En el entorno académico los dos perfiles eran equivalentes.

Finalmente, cruzamos los valores de las variables: contexto, competencias informacionales, actitud y perfil cognitivo. Los resultados obtenidos se muestran de forma en la tabla 7 para resaltar la relación entre variables. Es interesante destacar que en el entorno académico una parte notable de los identificados como no-expertos acostumbraban a tener un perfil de espectadores y caóticos. En cambio en el entorno profesional se identifica una relación directa entre ser identificado como experto en competencias informacionales y como implicado (actitud) y disciplinado (perfil cognitivo). Finalmente en el ámbito privado no se detectó que las competencias informacionales se relacionasen con los factores de actitud y perfil cognitivo.

3.7. Percepción en el uso de las TIC

Los resultados de esta variable mostraron que prácticamente no existían usuarios críticos en este estudio. Sólo en el entorno privado surgieron 3 casos. Pero en los ámbitos profesionales y académicos la gran mayoría eran entusiastas (17 y 19) y el resto eran realistas con el uso de las TIC.

Contexto	Competencias informacionales	Actitud / Perfil cognitivo			
		Disciplinado		Caótico	
		Implicado	Espectador	Implicado	Espectador
Académico	Experto	4	0	0	1
	No-experto	5	3	3	8
Profesional	Experto	11	3	1	1
	No-experto	2	4	0	2
Privado	Experto	6	1	6	2
	No-experto	0	3	1	5

Tabla 7. Resultados de la combinación de las variables: competencias informacionales, actitud, perfil cognitivo y contextos.

"Aplicando la combinación entre las variables actitud y perfil cognitivo se han obtenido 4 segmentos actitudinales: exhaustivo, pasivo, reactivo y proactivo"

3.8. Arquitectura de los segmentos actitudinales

Estos segmentos se obtuvieron a partir de la combinación de las variables actitud y perfil cognitivo en los diferentes contextos analizados.

En el ámbito académico, el cruce de variables condujo claramente hacia dos perfiles de usuarios. En este entorno, 9 participantes eran disciplinados-implicados y 9 eran caóticos-espectadores. A la primera combinación la denominamos "perfil exhaustivo" y a la segunda "perfil pasivo".

En el entorno profesional, también cruzando las variables actitud y perfil cognitivo, encontramos en primer lugar 11 participantes con un perfil exhaustivo, pero en segundo lugar surgió un nuevo perfil al que denominamos "reactivo" y que incluyó los 7 casos disciplinados-espectadores.

Finalmente, en el contexto privado, el perfil más numeroso fue el caótico-implicado (7 casos) que es lo que denominamos en este trabajo "perfil proactivo".

4. Discusión

Aplicando la combinación entre las variables actitud y perfil cognitivo se han obtenido 4 segmentos actitudinales que han sido denominados respectivamente de la siguiente manera: exhaustivo, pasivo, reactivo y proactivo. A continuación se detalla la definición y el perfil básico para cada una de estos 4 tipos de gestión de la información personal.

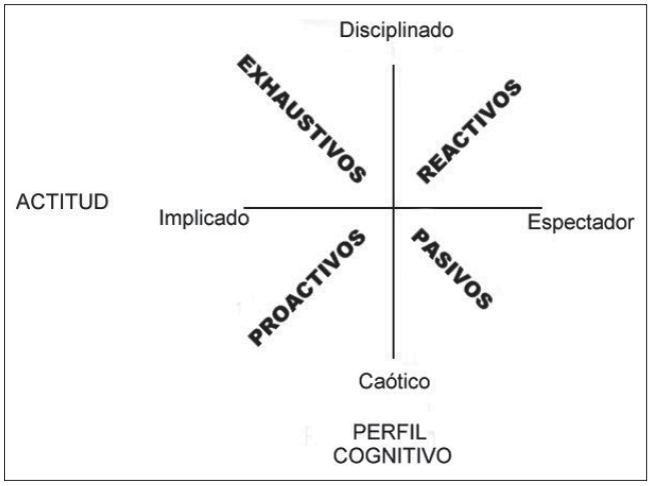


Figura 1. Los 4 perfiles de gestión de la información personal a partir de las variables actitud y perfil cognitivo.

4.1. Exhaustivo

Se corresponde con el perfil disciplinado-implicado. Son usuarios activos a la hora de buscar, gestionar y usar la información, resultándoles atractivo el descubrimiento de aplicaciones nuevas para optimizar los procesos. Cuando aprenden una técnica o descubren una nueva herramienta son muy hábiles a la hora de transferirlas y las utilizan simultáneamente en los tres contextos analizados. Esto justifica que su nivel de competencias informacionales sea alto y el mismo en todos sus ámbitos de actuación.

Los sujetos con este perfil utilizan aplicaciones actualizadas e innovadoras. En los entornos en los que se mueven necesitan estar al día para gestionar su información personal. Opiniones como "trabajo con bases de datos documentales digitales porque es más fácil gestionar la información" (del entorno profesional); "hago la web de mi pueblo, la actualizo y le voy añadiendo widgets" o "todo lo que hago está sincronizado con mi pda" (privado), apoyan esta idea.

Por otro lado, necesitan una hoja de ruta que les sirva de guía. Valoran de forma especial disponer de unas directrices pautadas exteriormente y del proceso para conseguir el resultado buscado. Afirmaciones como "el profesor ya me dice qué artículos necesitamos"; "en el trabajo no me han dicho cómo organizar los informes que hago, pero yo he ideado un sistema para generar un protocolo para futuras ocasiones", o "la música que me bajo de internet la tengo muy bien organizada por temas, tipo de música, grupos y cantantes" (privado), apuntan en esta dirección.

4.2. Pasivo

Se corresponde con el perfil caótico-espectador. Son usuarios inconstantes en los procedimientos y que tienen que adaptar sus reglas internas de funcionamiento a las reglas que rigen los contextos. Afirmaciones como "para organizar un viaje, por internet es más trabajo. Yo lo que hago es irme a la agencia y que me lo monten ellos" (privado), "no necesito organizarme, ya sé como están las cosas, es todo intuitivo" (personal), o "los programas que tenemos están muy estandarizados y cada vez que guardo algo se guarda solo, así que ya no tengo que hacer nada" (profesional), apoyan esta caracterización.

Tiende a asociar ciertas herramientas y técnicas a contextos determinados. Opiniones recogidas mediante las entrevistas como "me gusta más leer en papel que en pantalla" (académico) o "las fotos las tengo todas tal cual en digital en la carpeta Mis imágenes" (personal), van en esta dirección.

Los sujetos que presentan este perfil necesitan un entorno que permita disponer de una diversidad de vías

y procesos para llegar a resultados similares y que permita integrar herramientas que faciliten la gestión de la información y la optimización de procesos.

4.3. Reactivo

Se corresponde con el perfil disciplinado-espectador. Los sujetos con este perfil necesitan sentirse seguros con las aplicaciones y con los conocimientos para empezar a aplicar mejoras en la gestión de su información personal. Acostumbran a transferir procedimientos entre contextos de uso, pero respecto a las herramientas, tienden a utilizar en cada caso las que reciben de ese entorno o las que inicialmente empezaron a utilizar allí. Afirmaciones como "el *Excel* y el *Access* los utilizo más en el trabajo pero si quisiera también los podría utilizar en casa" o "El *Word* es para todos los documentos menos las presentaciones que se hacen en ppt" (profesional), apuntan en esta dirección. Las particularidades propias de cada entorno son las que les ayudan a cubrir las objetivos informacionales que persiguen: "la *UOC* me proporciona bastantes herramientas, así que, en principio no tengo por qué ir a buscar cosas fuera".

4.4. Proactivo

Se corresponde con el perfil caótico-implicado. Estos usuarios utilizan herramientas y las aplican en contextos diferentes, pero no con la intención de estandarizar sino como una inquietud personal. Afirmaciones como "al final tenía tantas cosas en favoritos que decidí hacerme una base de datos, porque no había manera de encontrar las cosas" (privado) o "es que internet y la informática forman parte de mi hobby, así que por eso invierto tiempo descubriendo qué puedo hacer de nuevo" (privado), apoyan esta caracterización. Suelen utilizar lo último en tecnología y buscan procesos que sean ágiles, flexibles y abiertos.

5. Conclusiones

Los 4 segmentos actitudinales presentados (exhaustivo, pasivo, reactivo y proactivo) pueden utilizarse como base para llevar a cabo proyectos centrados en los usuarios, que persigan la adquisición de competencias informacionales y la mejora de entornos virtuales.

Algunas observaciones destacadas en este trabajo ponen de relieve que hay una oportunidad para mejorar las competencias informacionales de los trabajadores y ciudadanos desde las universidades virtuales. El entorno académico puede estimular el proceso de búsqueda y uso de información teniendo en cuenta el nivel de competencias, la actitud y el estilo cognitivo de cada estudiante. Para ello es importante que las instituciones educativas eviten proporcionar de forma exhaustiva todos y cada uno de los materiales necesarios para el seguimiento de la asignatura y, en cambio, animen a los estudiantes a localizar, seleccionar, reutilizar, y compartir los recursos educativos. En este entorno el docente debe cambiar el rol de dispensador de conocimiento por el de facilitador de procesos de aprendizaje. El rol del estudiante pasa de ser un mero asimilador de productos educativos cerrados a ser un estudiante activo que adquiere competencias, conocimientos y habilidades.

Respecto al entorno virtual de aprendizaje, el académico puede ser, de nuevo, la locomotora de la introducción de herramientas y servicios y un destacado prescriptor de procesos. Para cubrir ese cometido debe contemplar aspectos relacionados con la personalización de esas herramientas y procesos teniendo en cuenta el estilo cognitivo, la actitud y las competencias informacionales de cada uno de los perfiles de los posibles usuarios.

"En un entorno virtual el docente debe cambiar el rol de dispensador de conocimiento por el de facilitador de procesos de aprendizaje"

Este trabajo de investigación forma parte de un proyecto de diseño centrado en el usuario de un repositorio de objetos de aprendizaje integrable en un entorno virtual personal (**Ferran** et al., 2009) que se encuentra en la fase de especificaciones del contexto de uso siguiendo los principios de la *ISO 13407*.

Nota

1. La traducción al castellano sería "estudio de incidencia" según **Tapia-Granados, José A.**; **Díez-Roux, Ana**; **Nieto, F. Javier**. "Gloepi: glosario inglés-español de términos de epidemiología y estadística sanitaria", *Boletín de la Oficina Sanitaria Panamericana*, 1994, v. 117, n. 3, pp. 239-257. http://www.ehu.es/PAT/Glosarios/epidemiologia.txt

Referencias

Dervin, Brenda; **Nilan, Michael S.** "Information needs and uses". *Annual review of information science and technology*, 1986, v. 21, pp. 3-33. ISSN 0066-4200.

Dörnyei, Zoltan. "Motivation and motivating in the foreign language classroom". *The modern language journal*, 1994, v. 78, n. 3, pp. 273-284. ISSN 0026-7902.

Ferran-Ferrer, Núria; **Guerrero-Roldán, Ana-Elena**; **Mor-Pera, Enric**; **Minguillón-Alfonso, Julià**. "User centered design of a learning object repository". En: Kurosu, M. (Ed.) *Human centered design*, *HCII 2009. Lecture notes in computer science* (LNCS) 5619, 2009, pp. 679–688. ISSN 0302-9743.

Jones, William. "Personal information management". *Annual review of information science and technology*, 2007, v. 41, pp. 453-504. ISSN 0066-4200.

Jones, William; **Teevan, Jaime**. *Personal information management*. Seattle, WA: University of Washington Press, 2007. ISBN 978-0-295-98737-8.

Katz, James E.; **Rice, Ronald E.** *Social consequences of internet use. Access, involvement and interaction*. Cambridge, MA: The MIT Press, 2002. ISBN 978-0262112697.

Krippendorf, Klaus. *Metodología de análisis de contenido. Teoría y práctica*. Barcelona: Paidos, 1990. ISBN 84-7509-627-1.

Léveillé, Yves. *Les six étapes d'une démarche de recherche d'information. http://pages.infinit.net/formanet/cs/chap2.html*

Neuendorf, Kimberly A. *The content analysis guidebook*. Thousand Oaks, CA: Sage Publications, 2002. ISBN 978-0761919780.

Palmer, Judith. "Scientists and information: II. Personal factors in information behaviour". *Journal of documentation*, 1991, v. 47, n. 3, pp. 254-275. ISSN 0022-0418

Richards, Lyn. *Using NVivo in qualitative research*. Bundoora, Victoria, Australia: Qualitative Solutions and Research, 1999. ISBN 0761965254.

Savolainen, Reijo. "Enthusiastic, realistic and critical. Discourses of internet use in the context of everyday life information seeking". *Information research*, 2004, v. 10, n. 1, paper 198. ISSN 1368-1613.

Talja, Sanna. "Analyzing qualitative interview data: the discourse analytic method". *Library and information science research*, 1999, v. 21, n. 4, pp. 459-477. **ISSN** 0740-818.

Virkus, Sirje. "Development of information-related competencies in European ODL institutions: senior managers' view". *New library world*, 2006, v. 107, n. 11/12, pp.467 - 481. ISSN 0307-4803.

Wilson, Thomas Daniel. "On user studies and information needs". *Journal of documentation*, 1981, v. 37, n. 1, pp. 3-15. ISSN 0032-0418.

***Núria Ferran-Ferrer**, Estudios de ciencias de la información y la comunicación, UOC, Rambla Poble Nou 156, Barcelona.*
nferranf@uoc.edu

***Mario Pérez-Montoro**, Facultat de Biblioteconomia i Documentació, UB, Melcior de Palau 140, Barcelona.*
perez-montoro@ub.edu