Project SEEKS

Initial Taxonomy of Information Seeking Behaviours

Deliverable 3

Mario Barajas and Elisabet Higueras

University of Barcelona

MINERVA SEEKS 90039-CP-1 **March 2003**

Index

1.	Introduction.	3
2.	Background.	3
3.	Research design.	9
3.1.	Methodology	9
3.2.	Data Analysis.	12
4.	Results..	13
4.1.	Preliminary case study.	13
4.2.	University of Barcelona SEEKS case study. . .	13
5.	Conclusions	17
6.	Bibliography.	20

Appendix:

A:	Code book for data analysis coding. . . .	21
B:	Preliminary questionnaire	23
C:	Scenarios	26
D:	Exit interview	27
E:	UB provisional proposal model. . . .	29
F:	UB Provisional taxonomy	30
G:	Data compilation charts.	33

1. Introduction

The SEEKS[1] project aims to develop a trans-nationally validated taxonomy of Information Seeking Behaviour to use as a tool for carrying out a census of the *Information Seeking Behaviours* (ISBs) of those groups of ICT users which have been less well-studied in previous ISB research. As a result, we may be able to indicate effective search strategy guidelines that can be used by educational software developers and ICT-based teaching and learning instructors for the production of appropriate curricula and training content. This effort will assist the inclusion of a broad group of users into full use of ICTs and thus into the Information and Learning Society.

We intend to effect this goal by observing ICT-user behaviour, specifically Information Seeking Behaviour. Within this work package, we will study previously defined target populations within different national sites in order to analyse which seeking strategies are used and then, develop a revised taxonomy, as one of the main outputs of SEEKS.

The objectives of this work package are:

- To review the existing taxonomies and build our own taxonomy and glossary (task 3) of the information seeking behaviours observed.
- To assist less knowledgeable users to carry over their stock of experience and expertise into the new context of ICT use.
- To indicate effective search strategy guidelines for educational software developers and ICT-based teaching and learning instructors to produce more appropriate and effective curricula and training contents.

2. Background

The research process followed is located within the Marchionini and Wilson models, described in the first deliverable[2]. The Wilson model is our starting point and has given us an initial framework, but its roots in information retrieval do present problems for us where the issue is user versus system for finding and organising information during searching. We are aware that in any human computer interaction there is a distance between the user's goals and knowledge and the level of description offered by the system: the distance between the user's thoughts and physical requirements of the system[3]

The overarching preoccupation of the SEEKS project is to begin now to prevent exclusion and we have located the aims of the project within the wider paradigm of, "Lifelong Learning". Zuboff (1988)[4] has stated: *'To put it simply, learning is the new form of labour'*. We take this to refer not merely to "Lifelong Learning,'" – as the need to acquire new skills and competencies in the context of a changing social, economic

[1] Information Seeking Strategies of adult learners in the Information Society.
[2] Research Review: a discussion framework Del 1.
[3] Hutchins et al., "*Direct Manipulation Interfaces"*,in User Centred Design System,ed.Norman et al
[4] P.395 cited in *Marchionini* 1995

and technological environment, but also to the necessity increasingly placed on citizens to maintain knowledge of markets, technology, law, health and safety information on a daily basis.
The following section will outline the previous research and sources used in our study.

Marchionini and Information Seeking

Marchionini (1995,9) situates Information Seeking within the context of *Learning Activities*. Within the wider context of Information Seeking, he identifies *Information Retrieval* (relative to a particular source) as breaking down into analytical strategies and browsing strategies. In the Table shown below, he suggests that these categories should not be seen as linear steps but as areas that remain in continuous interaction. He later maps these strategies onto a series of dichotomies:

Analytical Strategies	*Browsing Strategies*
Planned	Opportunistic
Goal Driven	Data Driven
Deterministic	Heuristics
Formal	Informal
Discrete	Continuous

In passing, we should note, that this does not mean that heuristics play no role in analytical strategies, but that they are more relevant before and after the execution, during planning and evaluation, whereas in browsing strategies they play a continuous role in guiding the process.

This author identifies the elements of Information Seeking Behaviours (ISBs) as the Problem (which itself is seen as arising from Needs), the Task, the System, and the Outcome. He makes clear that each set of these elements may or may not require access to more than one Domain. This raises the question of the transferability of ICT skills between domains, and whether obstacles to such transferability arise on the side of the System or on the side of the person[5].

As described in the first deliverable, Marchionini later draws the conclusion that domain expertise and system characteristics are more important for successful searching outcomes than ICT knowledge, suggesting that so long as the system is not a source of positive frustration, basic ICT moves are adequate for individuals who are able to autonomously evaluate the information provided and seek appropriate levels of information depth.

Marchionini (1995) reviewed different models of browsing and searching and he observed that, *"there seems to be agreement on three general types of browsing that may be differentiated by the object of search (the information needed) and by the systematicity of tactics used"* .

Directed browsing occurs when browsing is systematic, focused, and directed by a specific object or target: examples include scanning a list for a known item, and verifying information such as dates or other attributes.

[5] Marchionin and Komlodi (1998) attribute a system of four components –task, user, terminal, content – to Bennett (1972) while suggesting that "*most researchers would add a context component"*

Semi-directed browsing occurs when browsing is predictive or generally purposeful: the target is less definite and browsing is less systematic. An example is entering a single, general term into a database and casually examining the retrieved records.

Undirected browsing occurs when there is no real goal and very little focus: examples include flipping through a magazine and "Internet-surfing."

Marchionini (1995,49-60) proposes another often-cited model of the information-seeking process, tuned perhaps to electronic environments, and composed of eight sub-processes that develop in parallel:

(1) recognizing and accepting an information problem,
(2) defining and understanding the problem,
(3) choosing a search system,
(4) formulating a query,
(5) executing search,
(6) examining results,
(7) extracting information, and
(8) reflecting/iterating stopping

Wilson's Model of Information Seeking

In a report to the British Library Research and Innovation Centre, Wilson (1996) formulated a model of Information Seeking which encompassed the loop from information need to satisfaction. Wilson later (1997) identified the following categories of information seeking and acquisition after a survey of research that included health information seeking. As shown in his diagram (see appendix G), these information seeking behaviours will be the outcome of the entire model:

- *Passive attention*, such as listening to the radio or watching television programmes, where there may be no information-seeking intended, but where information acquisition may take place, nevertheless;
- *Passive search*, which seems like a contradiction in terms, signifies when someone finds information as an accident. E.g. watching info and you find is relevant for you.
- *Active search*, which is the type of search most commonly thought of in the information science literature, where an individual actively seeks out information; and,
- *Ongoing search*, where active searching has already established the basic framework of ideas, beliefs, values, or whatever, but where occasional continuing search is carried out to update or expand one's framework.

The "census" we plan to initiate will produce data on the relative presence of particular ISBs among different national user groups.

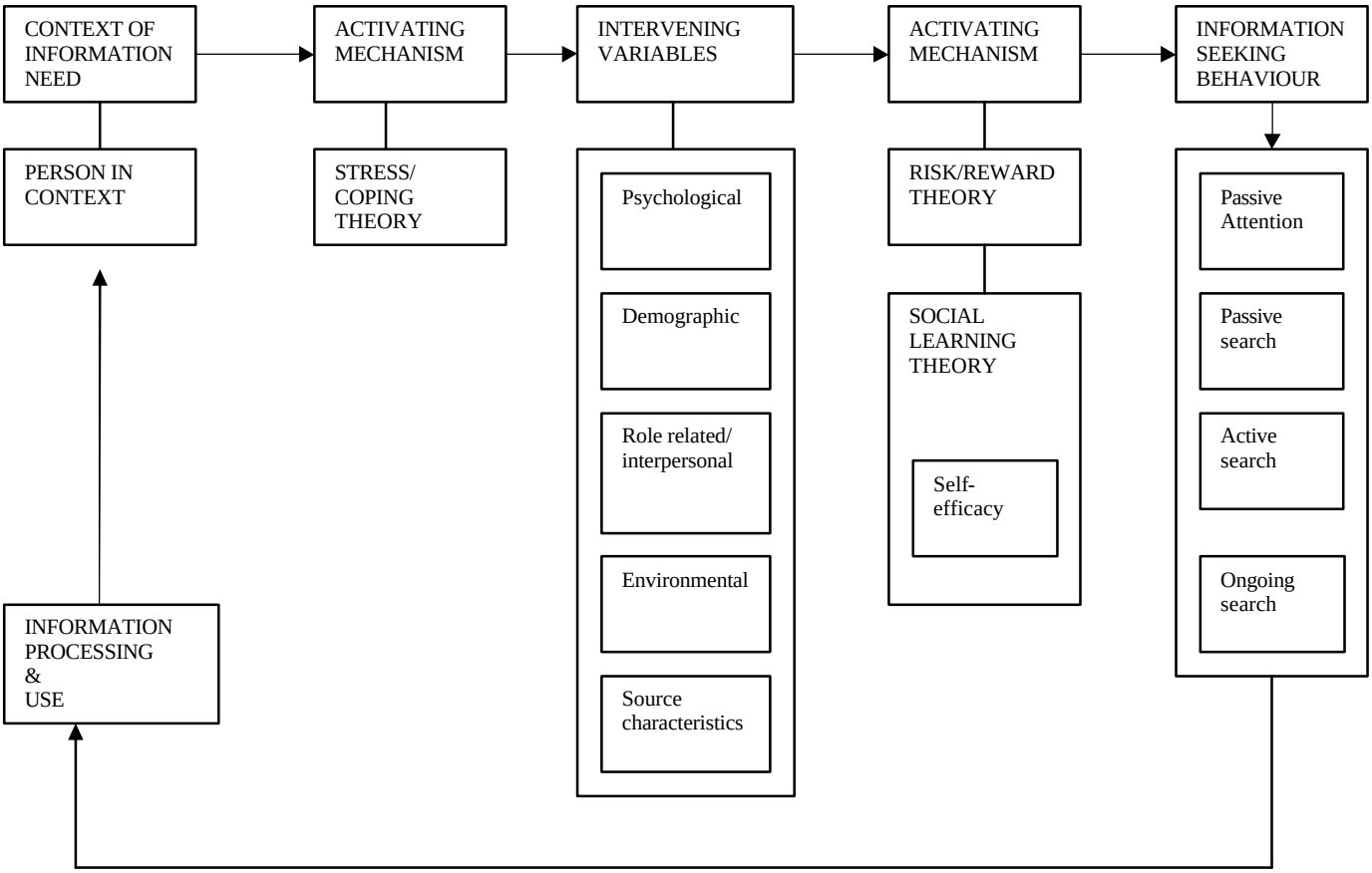
In Wilson's model, the final mix of ISBs is determined by a number of processes or contexts, which are the content of the first four columns of the loop:

- context of need;
- activation levels determined by stress/coping strategies;

- intervening variables of the situation;
- the feedback into activation levels from risk/reward factors in the resultant situation and arising from the learning process.

Within our proposed model, we will look for national differences which might be the result of different learning patrimonies, while attempting to correct for ephemeral factors, such as different levels of technological delivery systems. Intermediate factors might be the limited amount of cyber-space data available to smaller language communities.

In view of the limited time constraints on the project, it seemed worthwhile for us to adopt the latest version of the Wilson model as our basic framework model of IS and ISBs. We will be able to test this model and consider the need for amendments. The model is not claimed to be finished, so there is the possibility for addition of factors within the categories of intervening variables and ISBs. It will be a challenge, but, hopefully, a useful and fruitful one, to investigate possible national differences in the working out of the information need context, the stress and coping mechanisms, the risk/reward factors, and the learning theory aspects of the model. If the SEEKS project is unable to situate its findings within the context of this model, this will be a valuable outcome in itself.



Wilson's model diagram

Wilson uses an adaptation of Ellis' model which is outlined in the following section.

Ellis' Model of Information Seeking Behaviours:

Ellis (1989), Ellis et al. (1993), and Ellis and Haugan (1997) propose and elaborate a general model of information seeking behaviours based on studies of the information seeking patterns of social scientists, research physicists and chemists and engineers in an industrial firm. One version of the model describes six categories of information seeking activities as generic: starting, chaining, browsing, differentiating, monitoring, and extracting.

Starting comprises those activities that form the initial search for information - identifying sources of interest that could serve as starting points of the search. Identified sources often include familiar sources that have been used before as well as less familiar sources that are expected to provide relevant information. While searching the initial sources, these sources are likely to point to, suggest, or recommend additional sources or references.

Following up on these new leads from an initial source is the activity of *Chaining*. Chaining can be backward or forward. Backward chaining takes place when pointers or references from an initial source are followed, and is a well established routine of information seeking among scientists and researchers. In the reverse direction, forward chaining identifies and follows up on other sources that refer to an initial source or document. Although it can be an effective way of broadening a search, forward chaining is much less commonly used.

Having located sources and documents, *Browsing* is the activity of semi-directed search in areas of potential search. The individual often simplifies browsing by looking through tables of contents, lists of titles, subject headings, names of organizations or persons, abstracts and summaries, and so on. Browsing takes place in many situations in which related information has been grouped together according to subject affinity, as when the user views displays at an exhibition, or scans books on a shelf. ("Browsing" in the Ellis model is different from "viewing" in the previous section: browsing here describes looking for information at the micro-event level; whereas, viewing earlier describes a broader context of looking at information.)

During *Differentiating* activity, the individual filters and selects from among the sources scanned by noticing differences between the nature and quality of the information offered. For example, social scientists were found to prioritize sources and types of sources according to three main criteria: by substantive topic; by approach or perspective; and by level, quality, or type of treatment (Ellis, 1989). The differentiation process is likely to depend on the individual's prior or initial experiences with the sources, word-of-mouth recommendations from personal contacts, or reviews in published sources.

Monitoring is the activity of keeping abreast of developments in an area by regularly following particular sources. The individual monitors by concentrating on a small number of what are perceived to be core sources. Core sources vary between professional groups, but usually include both key personal contacts and publications.

Extracting is the activity of systematically working through a particular source or sources in order to identify material of interest. As a form of retrospective searching,

extracting may be achieved directly by consulting the source, or indirectly by looking through bibliographies, indexes, or online databases. Retrospective searching tends to be labour intensive, and is more likely when there is a need for comprehensive or historical information on a topic.

Ellis (1989) thought that hypertext-based systems would have the capabilities to implement functions indicated by his behavioural model. If we visualize the World Wide Web as a hyper linked information system distributed over numerous networks, most of the information seeking behaviour categories in Ellis' model are already being supported by capabilities available in common Web browser software. Thus, an individual could begin his or her information seeking by surfing the Web from one of a few favourite starting pages or sites (starting); follow hyper textual links to related information resources—in both backward and forward linking directions (chaining); scan the Web pages of the sources selected (browsing); bookmark useful sources for future reference and visits (differentiating); subscribe to e-mail based services that alert the user of new information or developments (monitoring); and search a particular source or site for all information on a particular topic (extracting). Plausible extensions of the activities to Web information seeking (labeled *Web Moves*), are compared with the original formulations (*Literature Search Moves*).

Taylor's scenarios of seeking information

Taylor (1991) compiled eight classes of information use generated by the needs perceived by groups of users in particular situations:

- *Enlightenment*: the desire for context information or ideas in order to make sense of a situation.
- *Problem understanding*: the need for better comprehension of the particular problems.
- *Instrumental*: The need to find out what to do/how to do something.
- *Factual*: the need for and consequent provision of precise data.
- *Confirmational*: the need to verify a piece of information.
- *Projective*: the need to be future oriented, concerned with estimates and probabilities.
- *Motivational*: the need to find additional information based on personal involvement with a task.
- *Personal or political*: the desire to control relationships, status, reputation, etc.

Wilson adapted those scenarios to his own model as context of information need (see diagram).

3. Research Design

The aim of this exploratory research, as we have indicated, is to develop a trans-nationally validated taxonomy of Information Seeking Behaviour, and to use it as a tool to carry out a census of the *Information Seeking Behaviours* (ISBs) of those groups of ICT users who have been less well-studied in previous ISB research.

The research comprises a series of pilot case studies (Yin, 1984) performed in different national sites. Data compilation techniques followed include: literature review (del.1), questionnaires, observation, individual semi-open interviews, and compilation of personal reports. Data analysis techniques are control and analysis of the variables, interpretation and contents analysis.

In order to provide a model for the partners to explore in their respective sites, the University of Barcelona undertook a case study which will be used as a referent for the other cases. The aim has been to provide a provisional taxonomy and a subsequent methodology which can be applied in the other cases. The partners will conduct a series of case studies at local level to determine what processes may be found to be in common within the cases when information is sought on the Web and to decide if the general ISB processes proposed by Wilson are applicable when using ICT- based tools.

The starting point is to provide preliminary evidence that supports a provisional taxonomy (*Passive engagement, selective searcher and dynamic searcher* (appendix G) that came out of a pilot study at the University of Barcelona. The proposed taxonomy might be transferred to other countries; nonetheless, SEEKS is conscious that any taxonomy can be dependent on cultural differences, demographic samples and other variables related to each country.

The aim of this preliminary study is to test the robustness and the transferability of the taxonomy and, if necessary, to reformulate a new one according to the trans-national results. (See appendix F & G)

3.1. METHODOLOGY

SEEKS will approach the exploratory research performing pilot case studies strategy, using such qualitative techniques as the following:

Profile of target groups

For this study the controlled target group that we propose is non-expertise adults IT users with one of the following profiles:

1. Participants in adult education
2. Tertiary level students in social sciences with low IT skills
3. Teachers in primary and secondary school with low IT skills
4. Women returnees to the labour market
5. Ethnic minority adults attending voluntary IT skills courses

It is suggested that partners adapt their participants to these categories, or justify why they are using different target populations. The target group can vary in number country-by-country. In any case, we propose to analyse groups that are no less than 5 users and no more than 15. The average should be a target group of 10. With these target groups, we expect to get enough people to define the population group we want to analyse, thus showing us the cultural and demographics difference to take into account in the taxonomy (Deliverable 4).

Scenarios

SEEKS is using different scenarios to solve a task responding to Wilson's definition of *Contexts of Information Need*. The scenarios are designed to solve these specific tasks:

1. FACTUAL SCENARIO, to find a *factual* answer to a specific question. This could be a yes/no answer or a specific fact that has been asked for.

2. LIST OF ALTERNATIVES SCENARIO, to assemble a *list of possible alternatives* for subsequent choice. This could involve lists of possible purchases, jobs, accommodation offers, holidays. SEEKS partners decided that this scenario would be common to all the countries.

3. INSTRUMENTAL SCENARIO, to assemble material to solve a problem.

4. CONFIRMATIONAL SCENARIO**,** to assemble material to support a case in a dispute.

5. MOTIVATIONAL SCENARIO, to acquire the essential background knowledge of a given field of knowledge. *(See Appendix C)*

This context of information need is based on Wilson's categories of *kinds of needs*. Wilson distinguishes three kinds of needs:

- new information
- elucidation of existing information
- confirmation of existing information

Each scenario is based on the kind of need that will guide the search task and the personal motivation, such as, Diversion, Personal relationships or Personal identity.
It is important, therefore, to take into account the *intervening variables* that Wilson lists as relevant:

- Personal
- Emotional
- Educational
- Demographic
- Social and interpersonal
- Environmental, economic
- Source characteristics (interfaces influence)

Based on the research framework SEEKS is interested in, a new variable has been proposed, which is important to take into account in our study in different countries:

- Cultural

This intervening variable will offer the possibility of testing the cultural differences among participating countries and populations.

Instruments for data collection

Entry questionnaires

The entry questionnaire is used to select the defined users among all the candidates. It is not of interest for SEEKS to investigate *experts,* so we prefer to use only the *experienced and beginner users* from the categories below:

> *Expert*: "I know how to find what I am looking for"
>
> *Experienced*: "I am quite good at finding information, but I could ask for help"
>
> *Experienced beginner*: "I can find information but I'm not very good"
>
> *Beginner*: "I can navigate the Web, but barely find what I am looking for"
>
> *Non Web User*: "I do not use the Web for finding information"

These questionnaires are given to a controlled target group with profiles defined as beginner, experienced beginner or experienced. In the mainly, the questionnaire will provide information on the independent variables (e.g. personal characteristics and learning aspects).

Observation

Direct Observation*:*
Moderated participation (Schatzman & Strauss, 1973), in order to get a "natural" environment where the participant can ask for elucidations forget the presence of the researcher. In this way the researcher can compile questions to ask at the exit interview.
Additionally, it will be data compilation of codes that were defined in order to compile the information that responds to Marchionini's processes (behaviours and their categories) of locating information, as a basis for initial categorization of the participants' information seeking behaviours (1995, pp.72-74): *patterns, strategies, tactics and moves*. (See Appendix A).
Indirect Observation*:*
Video recording, in order to contrast information gathered with other techniques and observe the body language. *Screencam* or log-files which might provide us with the process followed.

Interview

In the interview, the participant is asked how he/she has completed the tasks and what difficulties were found. It is important to ask the participant what skills an expert web user has. UB proposes that some basic questions might be asked in all the countries. The interview of approximately 10 minutes is of semi-structured, semi-open nature, so the observer can asks what s/he considers as relevant and avoid questions that are not important to a particular subject. Moreover, the interview gives information on

Intervening variables such as*:*

- *Personal characteristics*
- *Socio-cultural characteristics*: social perceptions of ICT using, auto-perception, expectations, context..
- *Professional characteristics*: applicability of Internet search, needs associated with the Internet search, satisfaction with the virtual environment, possibilities of improvement, personal use.

Data collection on Information Seeking Behaviours

To encode the behaviours, test applicability and completeness of taxonomy in case studies (Deliverable 3), we use additional codes related to the model proposed by Wilson:

- The scenarios proposed respond to the context information need that Wilson defines, and therefore to the kinds of needs related to.
- With the questionnaires, we control and take into account the variables.
- The qualitative observation gives information of the proposed pattern and what kind of cognitive behaviours are involved in the seeking process.
- The proposal of taxonomy is tested after the empirical research and reformulated in terms of trans-national transferability. (See Appendix C)

Collection of data for further analysis

In order to provide data to feed further analysis to be done on *Work Package 4 and 5* (specifically for the analysis of those software components and to provide guidelines for software producers), it might be necessary to compile and organise data related to the moves, tactics and strategies used by learners when using searching tools.

Marchionini's processes (Patterns – Strategies – Tactics – Moves) are used to analyse the observation and to categorise the collection data. With these items, we encode the most usual actions, and therefore we respond to the *SEEKS Deliverable 4* of providing guidelines to the software and web creators, in order to build up new software more adequate to the user's behaviours and usual movements.

The data collection will take into account the following actions:

MOVES**:** manifested as discrete behavioural actions
TACTICS**:** discrete intellectual choices or prompts manifested as behavioural actions during an information seeking session
STRATEGIES: sets of ordered tactics consciously selected, applied and monitored to solve an information problem; they can be general and flexible (browse strategies) or highly specialized and well-defined (analytical strategies)

Other kinds of information important for further analysis are the log-files and the screen-cam recording, in order to test the importance and the influence of the interfaces, what kind of websites were visited, and to do an analysis of this kind of information which is really intervening in the ISBs using the WWW. This is also an *intervening variable* inside the Wilson model (*characteristics of the sources*).

3.2.1. DATA ANALYSIS

Testing applicability and completeness of taxonomy

The aim of the experiment is to test applicability and completeness of the provisional taxonomy. After the data compilation, each partner will produce a report. An analysis of all the reports and their data will be made to provide a real transnational taxonomy.

Each report should detail changes and adaptations in methodology (as seen above), outcomes, and provisional national conclusions. The aim of the individual reports is to report data and outcomes and discuss possible generalisations and transferability.

4. University of Barcelona Results

A preliminary case study was developed and tested in Barcelona, in order to provide a framework for the other partners to perform their empirical study. The methodology employed was designed after a preliminary study done in a controlled setting, and will be described next.

The preliminary study was made in order to gain insights for the study, particularly regarding appropriateness of tools, the setting, the scenarios, and the methodology. The data of this preliminary study was not analysed as part of the SEEKS study because of incompleteness, although the data offered insights into ISB strategies. The methodology to be employed was designed after this preliminary study was done in a controlled setting, and will be described next.

4.1. Preliminary case study

Twenty students in the second course of the degree of Pedagogy were asked to do the test. They performed searches in eight scenarios and we aimed to test the best ones for the SEEKS study. This process was done in a group but as individual tasks. We concluded that:

1) For data compilation, it might be useful to add direct and indirect observation: observer and camera because of insufficient information provided by the personal reports.
2) There is a need to control the following variables: demographics, gender, Internet access, socio-economics, educational level, etc.

Using this preliminary design, the tools and techniques for data compilation selected to perform the SEEKS study lead to the following conclusions:

- Entry questionnaire: -necessary given the variability of the target population
- Scenarios: - eight scenarios will be tested according to the different information seeking needs
- Direct observation: - needed in order to enrich the data
- Personal report (open): -gave little information; students did not explain enough of their seeking processes.
- Exit interview: gives direct information about the decision making process in selecting the information

4.2. University of Barcelona SEEKS case study

Ten participants were asked to perform five scenarios (see appendix C) individually. In the preliminary study, the test was done in a classroom context. After data analysis, it was concluded that the classroom context was not a necessary factor. It was decided, therefore, to carry out the study on an individual basis. For further studies, the possibility of carrying out the study in a naturalistic context has been considered.

The study was conducted in December 2002, in a computer lab at UB. Each participant performed the five scenarios. The methodology used was the one described in the previous section. The observation model used was moderated participation (Schatzman & Strauss, 1973): In order to achieve a *"natural"* environment where the participant can ask for elucidations, s/he must be able to forget the presence of the researcher and the cameras.

Target population

UB was especially interested in looking at "agents of social inclusion": teachers, NGO volunteers, social workers, psychologists, pedagogues, etc.
The population of the study was a controlled and uniform sample of ten subjects (6 women and 4 men) with the profile shown below:

- Age: 20 – 30
- Experience in social inclusion work
- University degree

The reason for choosing young people is that this is the first generation to need the Internet and use it in their jobs. Consequently, the habits of use and the skills they demonstrate, would provide the study with factors that are important for inclusion.

RESULTS

a) Results provided by the observation

Type of Web site used

We can distinguish among three types of web sites used by the target population:

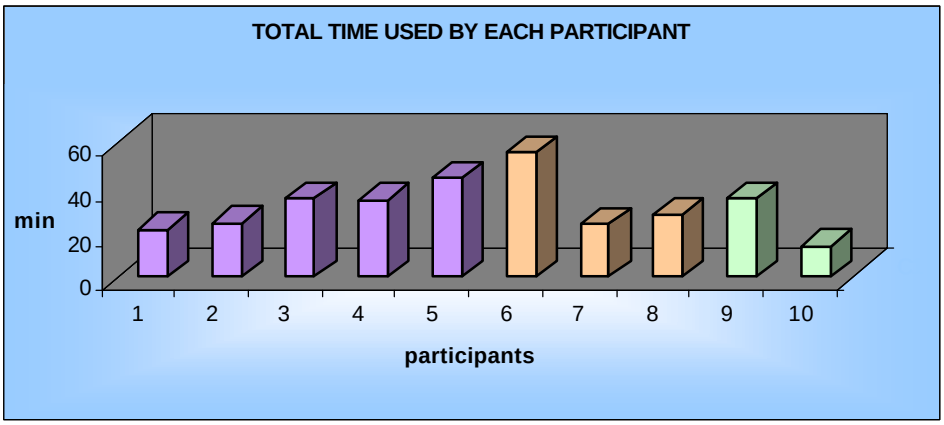
Commercial portal (in our case www.terra.es)
Search engine (in our case www.google.com)
Thematic website (many types depending on the subject)

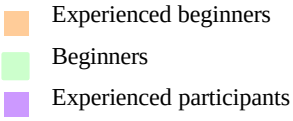
It was observed that:

Beginners only used commercial portals.
Experienced beginners used portals and search engines.
Experienced users used the three kinds of web sites.

Time spent

The time used by the participants was not relevant in performing the scenarios, but was for the quality of the information. Everyone spent more or less the same amount of time, but the experienced ones achieved the best quality results (in terms of completeness) in several scenarios:

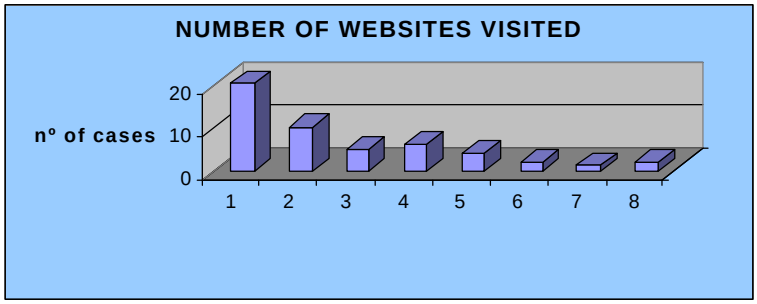




Number of alternatives

Only in 3 out of the 50 scenarios performed, did the participants (one different in each case) visit the second Web page of alternatives produced by the search engine. In no case did the participants check more than 8 websites, and in 20 cases out of the total 50 they only checked one website.

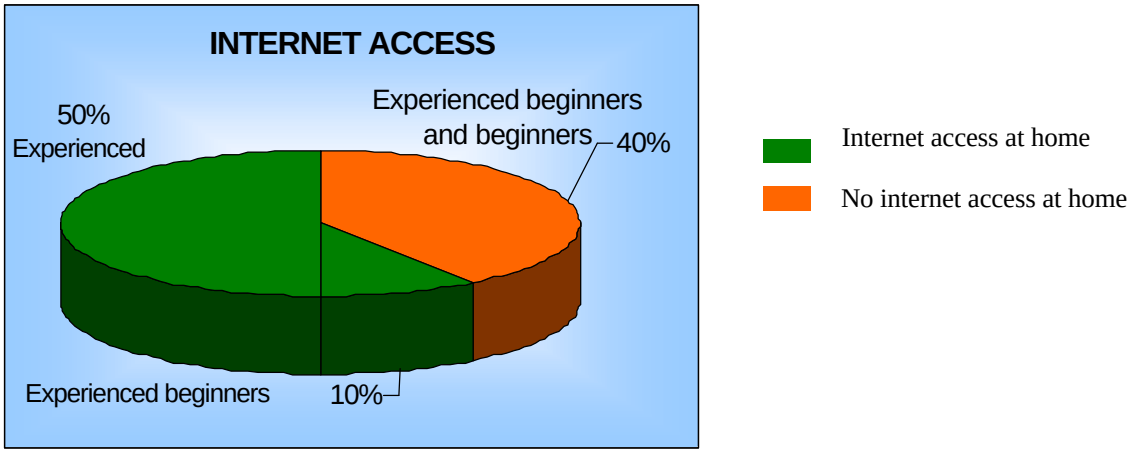
In the graphic below we can see the total rate of websites visited. If the query generated direct and relevant alternatives, then none of the participants visited more web sites.



b) Results of Questionnaires and exit interview

Internet Access-experience

The variable "*having internet access in home*" was of chief importance. Six out of the 10 participants had Internet access at home and 5 of them belonged to the experienced group. This result correlates with the experience in web use:



Another interesting aspect to analyse was the type of access they have at home (cable, ADSL, RDSI, modem, etc.), but in the UB case, all the participants having Internet connection had a modem, so this is an aspect to take into account in further research.

Preference of search engine

Thirty-eight out of the fifty cases used www.google.com. The main reason was related to usability understood as:

-Easy interface, very simple to use
-Speed of providing different alternatives
-Cited information in each alternative
-Automatic correction of errors
-Own language websites searching

The second reason was that the search engine used in the UB official Website is also www.google.com (6 of them had studied at the University of Barcelona).

Alternatives

Participants reported the following reasons for deciding on one website as relevant:

-Reliability
-Speed
-Quality of information
-Quality of the design
-Confidence
-Clear and simple format and language
-Own idiom
-Direct access to solve the need
-Non personal data required
-Previously known website or known from friends or media.

Gender

The study did not find any difference in terms of gender among the participants. This factor is going to be analysed in the following phase (Synthesis phase) after reviewing and synthesising all the case studies.

Correspondence with Wilson's model

Wilson's model is a general model of information seeking. The behaviours defined by this model are present also in seeking information in the Web, but not as the key behaviours found in this study, because of the nature of the Web environment. In any case, the relation found with Wilson's behaviours is the following:

Active search: The main behaviour while searching on the Web. It is present in all the scenarios observed. A Web search is always an active search.

Passive attention: This kind of behaviour is only relevant or visible in accessing known websites, e.g. a URL found in a magazine, advertisement, or newspaper, or recommended by a friend.

Ongoing search: Observable in people with previous knowledge in a subject; who seek information to elucidate their own background and/or to expand their framework.

Passive Search: No relevance in this study because it does not fit in. Wilson model's includes four categories but it is impossible to differentiate between passive search and passive attention.

5. CONCLUSIONS

Once the variables, the analysis of the results and the internal discussion were combined, the Barcelona team concluded that:

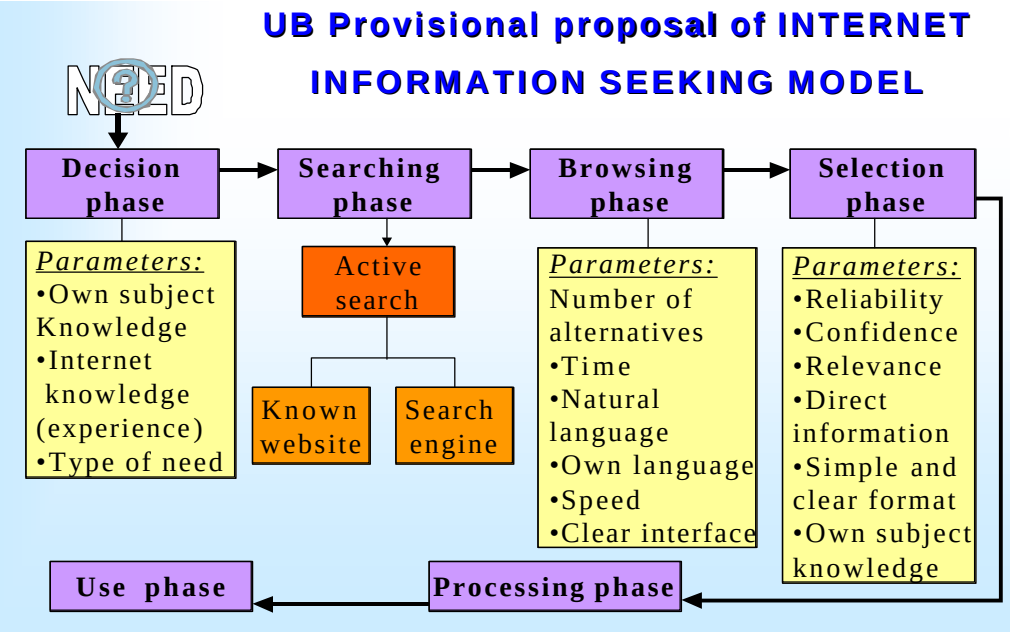
- There was not a *culture of using the Internet* for seeking information with a specific aim, yet.
- There were no criteria for selecting quality of the information on the net, and this factor influences the confidence of users with respect to the website. The quality and the design of the interfaces are of chief importance during the selection phase, and also for the final outcomes.

- There was not an association of internet use with every-day information seeking actions or most common queries. . For example, people who look for a mechanic or bank information will look to the Telephone Information Service, or call a friend or go directly to the site.

- People did not know how to select the best sources from all the information found. There was a lack of skills on how to select quality information.

- The process is a DECISION-making process in which the person decides how to search, decides what alternatives are relevant, and then SELECTS them.

Wilson's model is applicable to the Internet seeking process, but we have to know and differentiate the phases that characterize this kind of search. Wilson's model is more general and we have to adapt it to Internet searching. Here we have the phases identified as a preliminary model of seeking information using the Web:



As this was a quasi–experiment, it was not possible to analyse the phases of processing and information use, because the needs were not real personal needs. Depending on the parameters employed in the searching process, we can distinguish among three kind of seekers: Passive searcher, Selective searcher, Dynamic searcher). A correlation

between the behaviours and the Internet experience was identified: (Beginners, Experienced beginners, Experienced). As a general conclusion, Barcelona presents a taxonomy related to the experience in Internet use and the strategies that people use to search. This taxonomy came out of the empirical research and the data analysis done at UB.

The behaviours, low-skill IT users exhibit while searching the Internet, are present in the kind of strategy they follow. We distinguished three kind of users. The classification of this taxonomy is lower IT skilled user to higher IT skilled user, but we have only tested low IT users, so this taxonomy corresponds to this specific group:

a) PASSIVE SEARCHER

Low Internet use (corresponded with users who do not have home access). This kind of behaviour is correspondent with the strategies followed by the participants who were beginners:

DECISION PHASE: He/she does not differentiate among the types of need.

SEARCHING PHASE: Goes to a known website (e. g. a commercial portal).

BROWSING PHASE: When starting the *browsing phase,* only takes into account the parameters of:
- natural language(query),
- own language
- simplicity.

SELECTION PHASE: When the selection phase starts, the parameters which lead the search are:
- direct information,
- simple and clear interfaces.

If the information is not found, it is taken for granted that the information is not available on the Web.

b) SELECTIVE SEARCHER

Average internet use. The strategies used in this behaviour correspond to the users defined as experienced users.

DECISION PHASE: Type of need determines the searching phase.

SEARCHING PHASE: -Known website (a. Commercial portal) and Search engine (b.)

BROWSING PHASE: When starting the *browsing phase,* the following parameters taken into account:
- number of alternatives,
- natural language,
- own language,
- friendly navigation

SELECTION PHASE: In the selection phase the parameters are:
- direct information,
- clear format of information,
- subject knowledge.

c) DYNAMIC SEARCHER

Ample experience in Internet use allows the user different kinds of searching possibilities. The target population defined as experienced showed this kind of behaviour. The strategies followed to solve the scenarios were:

DECISION PHASE: Type of need determines the searching phase. Internet knowledge allows the access of different kinds of websites depending of the nature of the need. Having subject knowledge helps to solve the need faster and more satisfactorily.

SEARCHING PHASE: Obtaining more ways to decide where to start seeking:

a. Commercial portal
b. Search engine
c. Thematic website (specific for a subject)

BROWSING PHASE: When starting the *browsing phase,* he/she takes into account all the parameters defined in the UB model:

- Number of alternatives
- Time
- Natural language
- Own language
- Speed
- Clear interface

SELECTION PHASE: In Selection phase, all the parameters are present:

- Reliability
- Confidence
- Relevance
- Direct information
- Simple and clear format
- Own subject knowledge

The most experienced user behaviour achieved the best results, in terms of personal satisfaction. The knowledge of Web use comes from personal experience (no education about searching information and ICT), however, none of the three categories of users have the skills to do good selections. The information selection is the result of their experience and personal decisions.

In the next phase, this taxonomy should be confronted with the cultural differences of the participating countries in order to see similarities, differences, and the feasibility of transnational validity.

An interesting finding of the Barcelona study is the correlation between the Internet experience and the searching strategies. Also the strategies observed, provide information on the searching process and on the personal characteristics of the users.

6. BIBLIOGRAPHY

Cea, M.A. (1998). *Metodología cuantitativa. Estrategias y técnicas de investigación social.* Madrid: Síntesis Sociología.

Choo, Chun Wei, Detlor, Brian, & Turnbull, Don (1999). *Information Seeking on the Web - An Integrated Model of Browsing and Searching.* 1999 ASIS Annual Meeting Contributed Paper, and subsequently on firstmonday.

Jones, B. (2002) *Deliverable 1: Literature Review.* SEEKS – MINERVA. Manchester: MMU. (Internal document).

Kuhlthau, C.C. (1992) *Seeking meaning: a process approach to library and information services.* Norwood, New Jersey: Ablex Publishing.

Marchionini, G. (1995). *Information Seeking in Electronic Environments.* Cambridge: CUP.

Marchionini, G. & Komlodi., A. (1998). *Design of Interfaces for Information Seeking.* ARIST 33, pp. 89-130.

Monereo, C., Fuentes, M. & Sánchez, S. (2000) *Internet search and navigation strategies used by experts and beginners*. Barcelona: Interactive Educational Multimedia (pp.24-34). University Autonomous of Barcelona.

OECD (2001). *Education at a Glance – OECD indicators.* Centre for Educational Research and Innovation. Indicators of Educational Systems. Paris: Organisation for economic co-operation and development.

Polder, K.-J., Roeleveld, J. & Houtkoop, W. (in press). *Output indicators in vocational education.* Amsterdam: Max Goote Kenniscentrum, University of Amsterdam.

Spink, Amanda, Bateman, Judy, & Bernard J. Jansen (1998). *Searching heterogenous collections on the Web: behaviour of Excite users*. Information Research 4 part 2.

Spink, Amanda, Wilson, Tom, Ellis, David, & Ford, Nigel (1998). *Modeling User's Successive Searches in Digital Environments*. D-Lib Magazine April 1998.

Valles, M.S. (1998) *Técnicas cualitativas de investigación social. Reflexión metodológica y práctica profesional.* Madrid: Síntesis Sociología.

Walter, J.R. & Moen, W.R. (2001). *Identifying and categorizing Information Seeking Behaviors in the Networked Environment: An Exploratory Studyof Young adults. Final Report.* Texas: University of Texas.

Wilson, T., Ellis, D. Ford, N. & Foster, A. (1999). *Uncertainty in Information Seeking.* London: Library and Information Commission Research Report 59.

Wilson, T. & Walsh, C. (1996). *Information Behaviour: An Interdisciplinary Perspective.* London: British Library Research and Innovation Report 10.

APPENDIX A

CODE BOOK FOR DATA ANALYSIS CODING

To study the most usual movements while using the Web.

Processes:

MOVES		**TACTICS**		STRATEGIES	
M1	SCROLL	T1	REVIEW MATERIAL	S1	CHOOSING RESOURCE
M2	RETURN	T2	MODIFY QUERY	S2	KEYBOARD SEARCHING-INITIAL
M3	FRAME	T3	SWITCH RESOURCE	S3	KEYBOARD SEARCHING-SUBSEQUENTIAL
M4	TYPE ADRRESS	T4	COPY INFO TO ANOTHER SOURCE	S4	SUBJECT TREE SEARCHING
M5	USE OF ARROW			S5	LOCATING RESOURCE
M6	USE OF KEYBOARDS			S6	DECISION
M7	USE OF DROP DOWN MENUS			S7	VERIFY INFORMATION
M8	USE OF RIGHT CLICK YO OBTAIN/RETAIN INFORMATION			S8	CHOOSES SEARCH TOOL PREDETERMINED RESOURCE
M9	USE OF FORWARD KEY			S9	NATURAL LANGUAGE SEARCHING
M10	USE OF PRINT KEY			S10	USE PREVIOUSLY ACQUIRED INFO
M11	USE OF HOME BUTTON			S11	USE OF BOOLEAN CODES WITHIN SEARCH TOOLS

Definitions:

MOVES (Finely grained actions manifested as discrete behavioural actions):

M1: Scroll (using the scroll bar to look over the page)

M2: Return (going back to a previously viewed page)

M3: Frame (switching from frames to view or seek for information)

M4: Type address (type in the URL box)

M5: Use of arrow (use the mouse or arrows to scan or check for hyperlinks)

M6: Use of keyboard (to select or manipulate in a page)

M7: Use of drop down menus (select resources, seek information...)

M8: Use of right click to obtain/retain information (copy or paste)

M9: Use of forward key (going to previous viewed pages)

M10: Use the print function

M11: Use the Home button

TACTICS (Discrete intellectual choices or prompts manifested as behavioural actions during an information seeking session.):
T1: Review material (links to page and reviews material on screen to see if it is information needed)
T2: Modify Query (modifies query by changing or adding terminology in the search box).
T3: Switch Resource (switches to another type of resource)
T4: Copy Information from another resource (copy from one resource and uses it in another to try to locate answer).

STRATEGIES (Sets of ordered tactics consciously selected, applied and monitored to solve an information problem. Strategies can be general and flexible (browse strategies) or highly specialized and well-defined (analytical strategies). Strategies are the approach that an information seeker takes to a problem:
S1: Choosing resources (search tools, web pages...to begin)
S2: Keyword searching – broad (using the keyboard searching technique in the search query starting with broad or simple terms.
S3: Keyword searching – narrow (enters more specific keywords to narrow the search)
S4: Subject tree searching (seeking information using a subject tree technique in formulating terminology for the search query)
S5: Locating resources (tries to locate resources that will help solve the information problem)
S6: Decision (making a choice or selection regarding the information given)
S7: Verify information (checking to make sure the information given is correct or taht the participant understands the meaning of the task.
S8: Chooses Predetermined Resource (selects a resource to use from a list given by a search tool or web site)
S9: Uses natural language (forms thought into questions)
S10: Uses previously acquired information to continue to try and solve information problem
S11: Boolean Code Searching: boolean code operators within search tool's query box to create a search string.

APPENDIX B

PRELIMINARY QUESTIONNAIRE

UNIVERSITY OF BARCELONA Internet Questionnaire:

This questionnaire is the first part of a study to be conducted on how adults look for information on the internet. The following questions examine your abilities to find information and interests in using the World Wide Web.
From these completed surveys, we will choose the participants that meet the profile required for our study.

Remember to read the questions carefully and answer to the best of your knowledge:

✍ DEMOGRAPHICS (talk about yourself)

1. Name:
2. Age:
3. Status:
4. Gender:
   ☐ Male
   ☐ Female
5. Educational level:
6.
7. Job (if it's affirmative please tell what kind of job you do)
   ☐ Yes,
   ☐ No

8. If you work, what do you earn more or less:

☐Less than 350 € ☐between 350 – 700 € ☐between 700 – 1000€ ☐more than 1000€

9. Birth place and place of residence nowadays
   /
10. Who do you live with?_____
11. Do you have internet access at home?
    ☐ Yes
    ☐ No

12. Where do you use to connect to Internet?

✍ WEB KNOWLEDGE

13. A URL is :

☐ Uniform Resource Locators
☐ Web Address unique only to one Web Site
☐ Mailing server

14. What do gif, exe, zip, doc, html, & jpeg have in common?

☐ They help locate pictures
☐ They are all common types of files
☐ They are used in web addresses
☐ Do not know

15. What characteristics describe someone who can find information on the web really well?

16. How would you define your abilities to locate information on the web? (choose one)

☐ *Expert*: "I know how to find what I am looking for"

☐ *Experienced*: "I am quite good at finding information, but I could ask for help"

☐ *Experienced beginner*: "I can find information but I'm not very good"

☐ *Beginner*: "I can navigate the Web, but barely find what I am looking for"

☐ *Non Web User*: "I do not use the Web for finding information"

✍ WEB USE

17. Which browser do you use?
18. List three of your favourite search tools to find information:
    1)
    2)
    3)
19. What type of information do you like to search for (number in order 1=best to 11= least)

sports	entertainment	travel	food
games	music	news	cars
computer information	research for job		employment

20. How often do you find the information you are looking for? (circle the best that describes you)

☐All time ☐Most of the time ☐Sometimes ☐Not very often

✍ SEARCH TOOL KNOWLEDGE (What you know about the different search tools available on the web)

21. How do you locate information from the web? (choose all that apply)
    ☐ Use search engines
    ☐ Use subject trees
    ☐ Use pathfinders
    ☐ Use my own style of finding information(explain)_____

22. Have you used Google?
    ☐ yes
    ☐ no

22. Is it always better to use more keywords in a search connected by "and"?
    ☐ Yes, why_____
    ☐ No, why_____
    ☐ Do not know

23. In **Google** putting a "+" in front of keywords means (choose one):
    ☐ "and"
    ☐ "not"
    ☐ all hits may have the word
    ☐ some hits may have this word
    ☐ do not use **Google**
    ☐ do not know
24. In what order do most search tools list their results? (choose one)
    ☐ Chronological
    ☐ Alphabetical
    ☐ Frequency of term appearing
    ☐ Relevancy

25. Yahoo is a (choose one)
    ☐ subject search directory

☐ subject tree directory

☐ keyword search engine

☐ browser

Thank you for answering this survey. If you are selected, we will notify you when you will be required to perform some tasks using the World Wide Web. If you have any question please contact at:

Mario Barajas
University of BARCELONA
mbarajas@ariadna.d5.ub.es
or
Elisabet Higueras
University of Barcelona
ehigueal7@pedago.ub.edu

For easy contact with you, please complete the following information:
Name:
E-mail:
Phone number:
Cell phone number:
When can we contact with you? (morning, afternoon, night...)

APPENDIX C

UNIVERSITY OF BARCELONA INFORMATION SEEKING SCENARIOS

INFORMATION SEEKING SCENARIOS

1. Someone tells you that Paolo Freire has never studied pedagogy, you aren't agree otherwise you stay in doubt. Try to know if is it true or false. **(FACTUAL SCENARIO; *to find a factual answer to a specific question. This could be yes/no answer or a specific fact, which has been asked for*.)**

2. You want to go to New York this September and you don't know what's the cheapest agency to fly there. You neither know if is it more expensive to rent a room there or go with a organized travel with the hotel and flight included. You don't have too much money to spend so try to find the best solution to go to NY for a week. (describe the facts that make you choose one organized travel or your own searched travel) **(LIST OF POSSIBLE ALTERNATIVES SCENARIO; *to assemble a list of possible alternatives for subsequent choice. This could involve lists of possible purchases, jobs, accommodation offers, and holidays...)***

3. Tomorrow is your grandma's birthday. You know that she is a lover of the cuisine and you want to cook for her hers favourite cake: *tarta de santiago.* Try to find on the web the recipe of the cake and a picture to watch what is expected to be the last outcome of your cake. **(INSTRUMENTAL SCENARIO*; to assemble material to solve a problem)***

4. Someone has told you that in the film Oscar Wilde, the secondary actor is Jude Law. You think that is Ethan Hawke but you aren't sure because it was so many time ago that you watched the film. To finish the dispute you decide to look for it on the web. Who is mistaken? **(CONFIRMATIONAL SCENARIO; *to assemble material to support a case in a dispute*)**

5. You might know that Delors defined the four bases of the education. You read about this many years ago so you want to update your knowledge because a job partner has asked you to explain it to him. Find information of where there was written and what are this four bases of the education that are explained in the Delors' inform **(MOTIVATIONAL SCENARIO; *to acquire the essential background knowledge of a given field of knowledge*)**

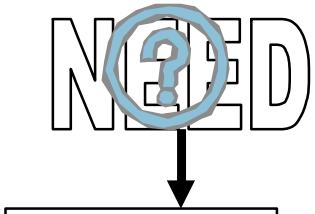
APPENDIX D:

University of Barcelona EXIT INTERVIEW QUESTIONS:

1. Where did you start your search in each scenario?
2. Why?
3. Why did you choose some information as relevant?
4. Are you satisfied with information ?
5. Do you usually compare information or do you take only the first one you get?
6. Why do you contrast some information and not others?
7. How do you know that some information is reliable or not?
8. Why do you choose some websites and reject others?
9. Do you typically start out on the Internet or would you start with something else?
10. In what scenarios won't you start searching by Internet?
11. Do you think that with more time you could make better searches or is it only a fact of Web knowledge what can improve your research process?
12. Is it important the characteristics of your Internet access (speed, public access...) to the quality of the research or to the experience?
13. Where do you usually look for this kind of information?
14. How you searched, is how you search when you ask someone? (natural language...)
15. How did you learn your skills to use Internet?
16. How could you describe someone who can find information easily on the Internet?
17. What characteristics do they have to become a good searcher?
18. Do you usually use different search tools as Google, Yahoo...?
19. Why do you use different ones?
20. Do you have to put in different terms for each different tool?
21. How do you learn about how to use those search tools?
22. How do you find out how best to use them?
23. Do you usually use webs that someone recommended you?
24. And , what other ways to know new websites do you know? (magazines, links...)
25. Do you ever go up to the help screen?

This is only a model but the researcher can ask anything that he/she thinks that would be of interest for the data compilation. The exit interviews are semi-open so instead this questions the researcher can ask anything he/she considers important in each case.

Appendix E: UB Proposal of INTERNET INFORMATION SEEKING MODEL



Decision phase

Parameters:

- Own subject Knowledge
- Internet knowledge (experience)
- Type of need

Searching phase

Active search

- Known website
- Search engine

Browsing phase

Parameters:
Number of alternatives

- Time
- Natural language
- Own language
- Speed
- Clear interface

Selection phase

Parameters:

- Reliability
- Confidence
- Relevance
- Direct information
- Simple and clear format
- Own subject knowledge

Processing phase

Use phase

Appendix F:

PROVISIONAL TAXONOMY

This taxonomy is result of the empirical research conducted by the University of Barcelona. All the partners are required to test it and, if necessary, in their reports they can propose and justify changes or a new one. The order of the taxonomy is less IT skill user to more IT skill user (lower to higher), but we have only tested low IT users, so this taxonomy corresponds to this specific group.

A. *PASSIVE SEARCHER*

Low Internet use (corresponds with users who do not have access at home). This searching behaviour is correspondent with the strategies followed by the participants who were beginners:

DECISION PHASE:
He/she doesn't differentiate among the types of need.

SEARCHING PHASE:
Goes to a known website (eg. a commercial portal).

BROWSING PHASE:
When starting the *browsing phase*, only takes into account the parameters of:

- natural language(query),
- own language
- simplicity.

SELECTION PHASE:
When the selection phase starts, the parameters which lead the search are:

- direct information,
- simple and clear interfaces.

If the information is not found ,it is taken for granted that the information is not available on the Web.

B. *SELECTIVE SEARCHER*

Average internet use. The strategies used in this behaviour correspond to the users defined as experienced users.

DECISION PHASE:

Type of need determines the searching phase.

SEARCHING PHASE:

- Known website (a. Commercial portal) and
- Search engine (b.)

BROWSING PHASE:

When starting the *browsing phase*, takes into account the following parameters:

- number of alternatives,
- natural language,
- own language,
- friendly navigation

SELECTION PHASE:

In the selection phase the parameters are:
direct information,

- clear format of information,
- subject knowledge.

C. *DYNAMIC SEARCHER*

Ample experience in Internet use allows different kinds of searching possibilities for the user. The target population defined as experienced demonstrated this kind of behaviour. The strategies followed to solve the scenarios were:

DECISION PHASE:

Type of need determines the searching phase. Internet knowledge allows access to different kinds of websites depending of the nature of the need. Having subject knowledge helps to solve the need faster and more satisfactorily.

SEARCHING PHASE:

More ways are available to decide where to start seeking:

- a. Commercial portal
- b. Search engine
- c. Thematic website (specific for a subject)

BROWSING PHASE:

When starting the *browsing phase,* he/she takes into account all the parameters defined in the UB model:

- Number of alternatives
- Time
- Natural language
- Own language
- Speed
- Clear interface

SELECTION PHASE:

In Selection phase, all the parameters are present:

- Reliability
- Confidence
- Relevance
- Direct information
- Simple and clear format
- Own subject knowledge

The most experienced user behaviour achieves the best results, in terms of personal satisfaction. The knowledge of Web use comes from personal experience (no education about searching information and ICT), however, none of the three categories of users have the skills to do good selections. The information selection is the result of their experience and personal decisions.

Appendix G: **Data compilation charts**

PARTICIPANTS CHART:

PARTICIP. 1	ENCODED SCENARIOS				
	1	**2**	**3**	**4**	**5**
URL initial					
Nº websites visited					
Nº queries					
Most common actions					
Other URLs typed					
Use of local website menu					
Information contrast					
Looks in other results pages					
Observations					

Definitions:
URL initial: write the website where the participant starts the search. Ex. www.google.com
Nº websites visited: write the number of websites watching the screen-cam. Ex. 4 or 5....
Nº of queries: how many queries write the participant in the searching box.
Most common actions: browsing through the results, scroll down. Check different options, do queries, or what the observed decides as the most common.
Other URLs typed: if the participants goes to a known website (not by linking) then write the websites visited, ex: www.easyjet.com, www.rumbo.es , etc.
Use of local website menu: yes or not. If the participant searches by the website menu write yes, if only uses the search engine write no. If uses both write yes.
Information contrast: if the participant contrasts the information then write yes, it is useful in this case check the exit interview in order to know why (satisfaction, relevancy, not clear interface...).
Looks in other results pages: if using a search engine he/she browses the second or others pages of results, as in google pass to the second page of results.
Observations: information that the researcher wants to observe more exhaustively or not usual.

SCENARIOS CHART

SCENARIO 1	ENCODED PARTICIPANTS									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
URL initial										
Nº websites visited										
Nº queries										
Most common actions										
Other URLs typed										
Use of local website menu										
Information contrast										
Looks in other results pages										
Observations										

The same information but like this you can analyse per scenario and not exclusively per participant.

INDEPENDENT VARIABLES CHART:

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
AGE										
Gender										
Degree										
Field degree or job										
Home internet access										
Web knowledge										
Salary										