A coding system for qualitative studies of the information-seeking process in computer science research

Cristian Moral, **Angelica de Antonio**, **Xavier Ferre** and **Graciela Lara**
Escuela Técnica Superior de Ingenieros Informáticos, Universidad Politécnica de Madrid, Spain

Introduction. *In this article we propose a qualitative analysis tool –a coding system - that can support the formalisation of the information-seeking process in a specific field: research in computer science.*
Method. *In order to elaborate the coding system, we have conducted a set of qualitative studies, more specifically a focus group and some individual interviews, and have analysed their results.*
Analysis. *We provide a detailed description of how the qualitative studies were performed and how the subsequent analysis, refining and validation phases were carried out until we obtained the proposed coding system.*
Results. *The coding system is presented as a list of hierarchically categorised codes that try to cover the full information-seeking process performed by researchers in computer science. In order to facilitate the understanding and later use of the codes, we also include a detailed description of each, together with real examples of use extracted from the qualitative studies.*
Conclusions. *We present a complete list of 169 codes categorised into 13 categories that can help other researchers to analyse further qualitative results related to the information-seeking process, both at the process level and at the user level. In our case, these codes will be in the near future the seed to design a system that can adapt its functionality, visualisation and interaction based on the task at hand, the user and the context of use.*

Introduction

Seeking information is a key activity in research that consists of two main components: search and exploration. Search refers to the tasks performed in order to obtain a first corpus of results. Some activities related to this component are query construction or selection of the information source where the search is going to be performed. On the other hand, exploration refers to the

tasks oriented to determine which of the preliminary results are actually interesting and useful to meet the information needs the researcher has at that time. It is also related to the habits and standards adopted by the researchers to organize and classify the selected information.

In every field of study, researchers are required at some point to look for information. However, information-seeking is still one of the most difficult and time-consuming activities in research. Even if there are many generic and specific search engines, the extremely large amount of information now available and the high range of possibilities for information use make it very difficult to standardise and generalise the process. The information-seeking process, especially the exploration component, requires from the user an effort to individually process the collected information. Individual characteristics of the user, like his/her professional profile, preferences, capabilities, or experience have a great influence on how the process develops. Additionally, many other external factors can affect the process, like the purpose of the search, the context of use - in space and time - or the available technology. Therefore, a one-size-fits-all approach is not the best solution, as all the differences, similarities and specificities are then inevitably ignored. Today there are a lot of tools – search engines, meta-search engines, scientific repositories, reference managers, impact metrics... - but most of them do not take into consideration and adapt to the singularities of the user and the context wherein the information is sought. The goal of our research is to demonstrate that the information-seeking process is not universal and consequently support tools need to be adaptive and adaptable, so that the system moulds itself to the user, rather than the user moulding to the system.

To achieve this goal, the information-seeking process had to be studied and analysed in order understand it more perfectly. We believed that a deeper understanding would allow us to develop a reference model of the information-seeking process performed by researchers in computer science that takes into account the differences, similarities, and singularities existing for different users in different conditions and situations.

We have run a set of qualitative studies in order to elaborate our theory based on real data from real researchers. The use of rigorous tools is needed to properly analyse qualitative data, which is essential to ensure the correctness of the model. In this paper we present one of these tools, a coding system, that has emerged from the analysis of the data. The coding system is presented as list of codes that have been hierarchically grouped and ordered according to the concepts they are related to. A complete definition of each code has been included, together with real examples of use in coding system individual interviews. In the future, the concepts that emerged will be the seed for the construction of a model of the information-seeking process performed by researchers in computer science. This model, in turn, should inform the design of adaptive and adaptable support tools.

In the remainder of the paper, we first review the general methodology we followed, while in sections 3 and 4 we detail how the two qualitative studies – the focus group and individual interviews, respectively – were carried out. The description of how the obtained data were analysed and how the coding system emerged from these data is provided in Section 5. Final results after the refinement and validation of the coding system are presented in section 6. Finally, some conclusions and descriptions of future work are given in Section 7.

Methodology

The information-seeking process is a very familiar activity for researchers, but it is difficult to formalise. Our first approach was to informally ask some of our colleagues how they seek for information during their research activity. Some of them were unable to explain how they look for, select, classify and archive papers for their research. Either they had interiorised and mechanised the process to such a degree that they did not remember how they usually perform

it, or they considered the process un-quantifiable because it depends on many external factors. There were some of them who attempted to propose a definition of the process, but there was a high variability between them. In both cases, the results would have probably been wrong, incomplete or forced.
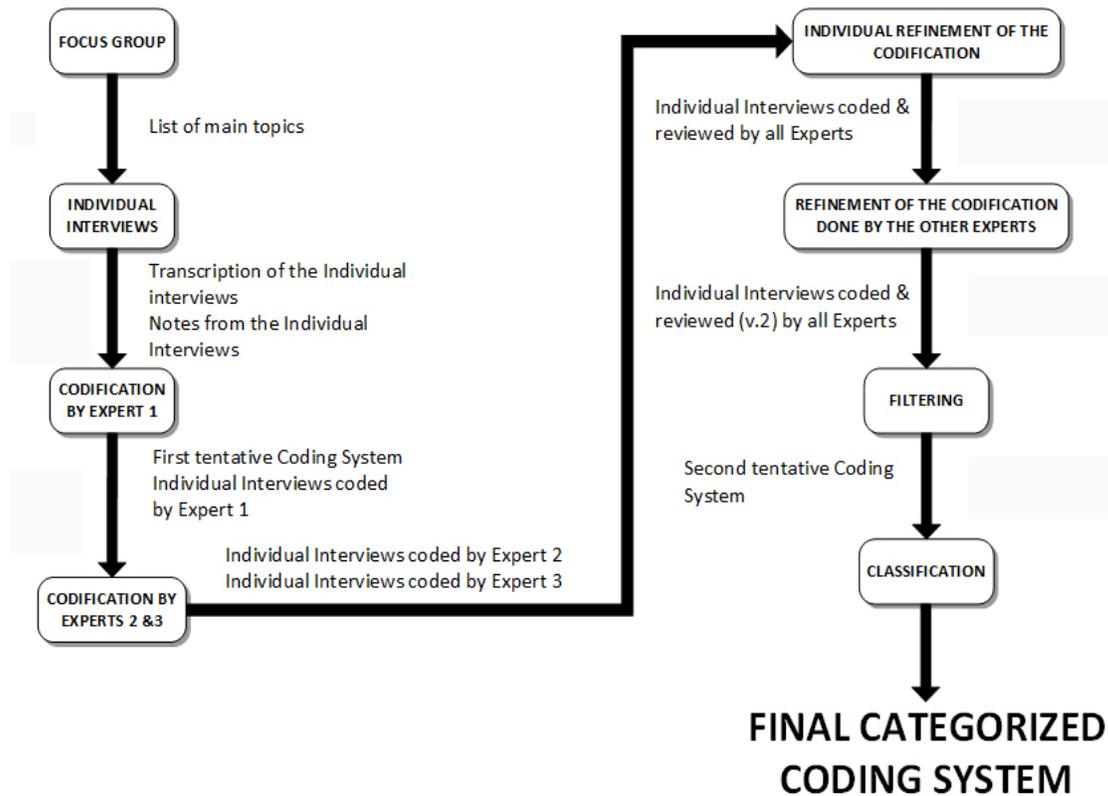

Figure 1: Steps followed to carry out the qualitative studies

In order to overcome this lack of definition and the potential subjectivity, we ran a set of qualitative studies (see Figure 1) so as to increase our understanding of the process and objectivise the process as much as possible.

First of all, we decided to run a focus group (Krueger and Casey, 2009) formed with researchers, in order to provide them with a relaxed environment where they could share and contrast their ideas and opinions. In this way, researchers who initially were unable to define the process could discover similarities with other researchers so that they could better realise how they seek for information. In turn, the discussion would help identifying the similarities and differences in the process between them.

At this point, the main aspects and ideas stemming from the focus group were the starting point to develop a questionnaire that served as a guide for running individual interviews to researchers. Basing the questionnaire on a previous group interview with researchers decreased the risk that important ideas might be omitted in the interview design. These interviews were the main source of data collection for the formalisation of the information-seeking process in the context of computer science research.

It is important to highlight that both the focus group and the individual interviews were conducted in Spanish, and then the data were also collected in Spanish. In order to facilitate and internationalise the results of our study, we have reliably and thoroughly translated both the tools, namely guides, questions, and keywords, used during the qualitative studies and the results obtained from the qualitative analysis, ensuring that both the content and the meaning were the same.

During the whole process we established as an essential and leading criterion the achievement of a high level of quality in both the design and the conduct of the qualitative studies, and in the obtained results. To reach this goal, we used as a guideline the criteria proposed by Tong, Sainsbury and Craig (2007) when reporting the results that emerged from focus groups or interviews. Explicit and detailed information has been provided in this paper about the personal characteristics of the members of the research team and their relationship with the participants; the selection of the participants; how and where the data were collected; how the data were analysed and how the results were reported. We also considered some issues that, according to Miles and Huberman (1994), should be addressed both during the design and conduct of qualitative studies, and when reporting the results, in order to ensure the quality of the obtained results. To be more specific, we took into consideration the *confirmability of the results* – e.g., the methods and procedures we followed are explained in detail –, the *reliability of the results* – e.g., the research question has been made clear, and the researchers' role and status have been explicitly described –, the *internal validity of the results* – e.g., results have been triangulated –, the *external validity of the results* – e.g., details of the sample have been provided and we have defined the boundaries and scope of the results for generalisation – and the *application of the results* – e.g., the results are available to potential future users and they potentially can help to answer the research question posed.

Focus group

Research team

As the number of participants in the focus group was relatively large, the study was led jointly by two members of the research team.

The second author of this article was the facilitator, whose role is to guide the discussion in order to ensure that participants do not deviate from the topic. To do this, the facilitator can pose some general questions to launch the conversation or to go deeper into an interesting thought or idea mentioned by one or more participants (Kitzinger, 1994). This does not mean that the facilitator poses questions and the participants just answer them, but rather the opposite. Participants have a high degree of freedom to redirect the discussion to any topic they consider interesting. The facilitator's role is to ensure that this deviation still keeps the discussion on track. The second essential aim of the facilitator is to manage the speaking time of the participants, so everybody can express his/her opinion. This is extremely important in order to avoid having answers only from dominant participants, to the detriment of shy or low confidence participants (Kitzinger, 1995). Additionally, results are richer if all the participants give their opinion and participate in the discussion. In our case, the role of facilitator was undertaken by a senior researcher, a full professor at the University, female and aged between 45 and 50. The facilitator has a large experience running qualitative studies, which was considered essential to ensure the proper development of the focus group.

To give support to the facilitator, the first author took the role of notetaker. The purpose of the notetaker is to annotate things that go beyond the speech itself, items of interest that can provide more details to the words of the interviewees (Powell and Single, 1996). Into this category falls the so-called non-verbal communication - the tone and volume of the voice, or the gestures and postures the speakers make during speech, among others. In this case, the role of notetaker was played by a Ph.D. student in computer science, male and with no previous experience running qualitative studies. The notetaker remained silent during the session in order not to break the group dynamics and to avoid interfering in the facilitator's work.

Research sample

The research sample was composed of 9 participants. The selection was performed under a set of strict restrictions in order to obtain not only relevant, but also significant results.

All participants were required to have a professional background in the same field under study. This meant that all participants, one way or another, were involved in some kind of research in the field of computer science. We tried to consider as many profiles as possible, and then selected both senior and junior researchers in the field. The first group was represented by five university professors, all of them with a Ph.D. in computer science, while the second one was formed by four Ph.D. students in computer science.

As computer science can be divided into many specific areas, we also tried to cover most of these areas in the focus group. Each of the professors represented a different department inside the computer science School (UPM): [Computer Systems Architecture and Technology](), [Artificial Intelligence](), [Linguistics Applied to Science and Technology](), [Computer Languages and Systems and Software Engineering](), and [Applied Mathematics](). In turn, the Ph.D. students were researching a variety of topics - Software Engineering, Virtual Environments, Data Mining and Intelligent Software Agents. In demographic terms, participants were aged from 28 to 50 years old, with three female and six male.

All participants were contacted via email. The message was standard for all the candidates and explained in general terms the purpose of the focus group and how it would take place. It also informed potential participants that there would not be any kind of compensation or incentive, apart from helping us in our research topic. At the end of the email, they were requested to indicate their interest in participating in the study. The ones that answered affirmatively were replied to with a message of gratitude and a request to choose the most suitable date for them to participate.

After selecting the most popular date, one of the participants had to cancel her participation because of incompatibility with the schedule, but she proposed another participant with a similar profile in terms of field of research, age and level of expertise as researcher.

All the participants - except one Ph.D. student - and the team members running the study knew each other to some extent, either as coworkers, from membership in the same research lab, or for having attended some course taught by another participant.

Study design

Introduction

The facilitator welcomed everybody and thanked them for their participation. In order to provide some context, she reviewed the origin and purpose of our research and explained, in general terms, the objectives and the procedure of a focus group.

As a gold standard for the whole session, the facilitator asked all the participants to be as sincere as possible. She made it clear that no answer would be considered as a bad answer but, on the contrary, all answers would be good, interesting and relevant for our research. Finally she stated very clearly that in any case, none of them or their methods would be questioned, and then asked them to feel comfortable, recognizing that there was no pressure.

Open-ended questions

The discussion started with the following question: *'How do you seek for information while carrying out research?'* During the nearly two hours of the session, the facilitator intervened

very few times to help participants to remain focused, or to motivate some of them to participate.

The facilitator used a set of keywords previously defined by the research team as a guide. These keywords referred to general concepts that were considered potentially related to the topic. Some examples are phases, objectives, tasks, context, user profile, tools, workspace, search, filtering, classification, visualisation or interaction. With these keywords, we expected to obtain answers to questions like: *Which are the steps of an information-seeking process? Which are the problems and deficiencies you usually have to face while looking for information? What else would you like a system to do to help you during the information-seeking process?* However, no predefined questions were used, to allow the discussion to flow and to avoid prejudicing the participants' opinions.

Setting

The session took place in a meeting room of the computer science school, which is the common workplace of both the participants and the research team. Participants, facilitator and notetaker sat at a round table in order to facilitate the communication and provide a feeling of equality, so as to promote participation. As the focus group was expected to finish just before lunch time, some finger food and beverages were provided during the discussion to help participants feel comfortable and focused.

Data collection

In order to allow both the facilitator and the notetaker to be focused on the discussion, the session was audio-recorded using a smartphone. Before starting the session, all participants were informed of the recording and were asked for permission. The facilitator also informed the group that the recording would be only used to analyse the discussion and that, in no case would individual opinions would be identified in any published report or subsequent work.

No further interviews were made with the participants to corroborate their discussion and the final transcripts were not returned to the participants. We did not ask for feedback from the participants on the findings.
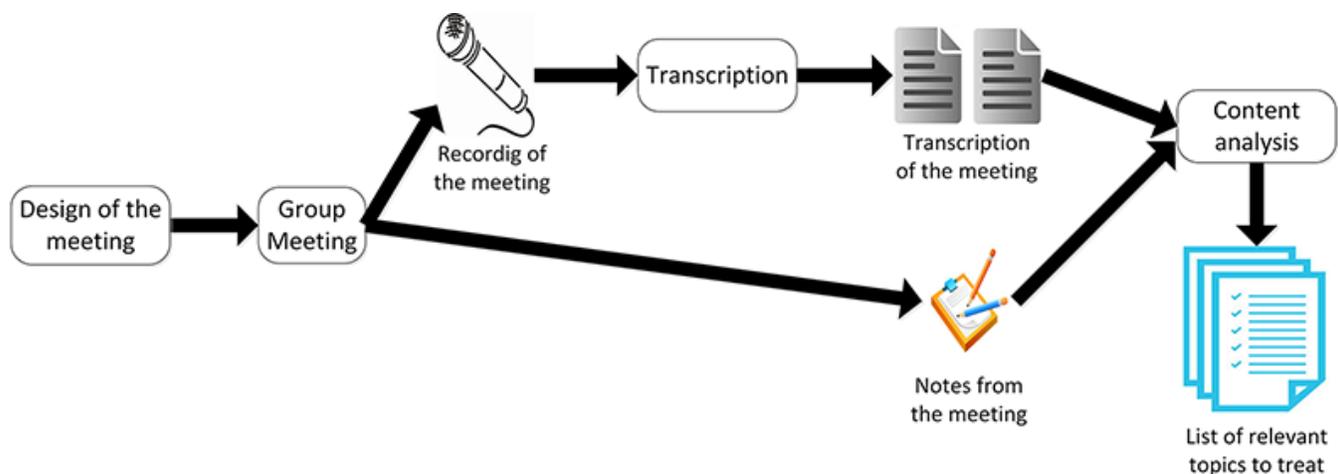
Data Analysis

Figure 2: Focus Group

Once the session was transcribed, the research team analysed and discussed the content of the focus group (see Figure 2).

The intent was to identify which aspects are involved in the information-seeking process. As a result, the analysis consisted of grouping related ideas or comments so as to draw a preliminary conceptual map of the main themes that should be considered while studying the information-seeking process. At the end, 13 concepts related to the information-seeking process were identified:

1. Archiving
   Comments related to the activity of archiving of the articles during the information-seeking process.
2. Corpus
   Comments related to the set of results obtained after a search process.
3. Search process
   Comments related to the process of looking for results that may cover an information need.
4. Reading process
   Comments related to the process of reading the articles selected after the information-seeking process.
5. Exploration process
   Comments related to the process of exploration of the results obtained from the search process.
6. Wishes
   Comments related to things the user would like to happen during the information-seeking process.
7. Difficulties
   Comments related to difficulties found during the information-seeking process.
8. Collaborative work
   Comments related to the information-seeking process performed collaboratively in a team.
9. Workspace
   Comments related to the organization and use of the physical and digital space during the information-seeking process.
10. Tasks
    Comments related to tasks undertaken during the information-seeking process.
11. Meta-information
    Comments related to meta-information used during the information-seeking process.
12. User profile
    Comments related to the profile of the user with respect to the information-seeking process.
13. Reference manager
    Comments related to reference managers used during the information-seeking process.

Individual interview

In order to refine the initial set of concepts obtained from the focus group, we carried out a set of interviews with the aim of obtaining a set of adequate codes for tagging information-seeking activities in this computer science research. The results of the focus group served as a basis for the design of the questionnaire. The interviews were conducted individually in order to allow each participant to explain in detail how he/she seeks for information while researching.

Before running the interviews, we developed a semi-structured questionnaire to guide the sessions. Some of the questions were open-ended in order to allow wide and unbounded answers from interviewees, but also in order to avoid biasing the participants by unintentionally leading their answers. As an example, the first question *'How do you seek for scientific papers while doing research?'* was very general and open-ended, allowing the user to express almost

anything related to the studied process. Some more specific questions were also posed when we wanted to obtain more details about more concrete aspects, but even in these cases the interviewees were able to express their ideas openly as they were asked to explain more in detail their answers: *'Do you think the information-seeking process varies according to the context? Please explain your answer'*. Additionally, the interviewer had the freedom to pose new questions to delve into interesting comments, or to clarify incomplete or unclear answers, thus making the interview more flexible, in order to obtain as much information from the interviewees as possible (Bernard, 1988; Crabtree and Miller, 1999).

Research team

In order to maintain the consistency in the interviewing process, all the interviews were performed by the same member of the research team. In this case, the notetaker of the focus group was considered the best member to carry out the interviews, as he had been present during the group interview.

Research sample

The criteria followed to recruit participants were exactly the same as in the focus group: senior and junior researchers from different areas of the computer science field, covering a range of ages - in this case from 27 to 55 years old. In order to avoid interviewing biased participants, all the researchers that participated in the focus group were considered ineligible for this second qualitative study. Again, potential participants were contacted via email using a standard message, explaining to them the purpose of the interview and how it would take place, and informing them that participation was unpaid. Among those who answered the email indicating they were interested in participating, 5 females and 3 males were selected.

Once again, the interviewer and the interviewees knew each other to some extent, either because the interviewer had previously attended some course taught by the interviewee, or because both worked in the same research lab.

Study design

Introduction

The interviewee was first welcomed and thanked for his/her participation, and then the interviewer introduced him/her to the purpose of our research. Details were provided as to how the interview would proceed.

The interviewer also asked the participants to be as sincere as possible, stating clearly that we were not evaluating neither him/her nor his/her methods, but only trying to figure out how the information-seeking process is carried out from different perspectives.

Open-ended questions

Even if predefined questions were used to guide the interview, the interviewee was allowed and encouraged to explore topics, reflect, and generally express everything he/she considered relevant for the question or for any other topic he/she considered related and/or interesting. Our main goal was to bring to the surface all the concepts, problems, suggestions, difficulties, tasks, and so on, related to his/her information-seeking processes. As a result, we let the interviewee develop his/her ideas and thoughts as much as he/she wanted.

To reinforce this goal, the interviewer was allowed to pose unplanned questions not present in

the script when needed in order to dig deeper into an idea, to clarify an ambiguous answer, or to nail down an interesting comment.

Setting

The location and schedule for the interviews was agreed individually with each of the participants according to their availability and preferences. Five weeks were needed to perform all the interviews, and seven of the eight interviews took place in the interviewee's workplace, while one of them was carried out in the interviewer's workplace.

Only the interviewer and the interviewee were present during the interview to avoid external influences that could disturb the participant.

Data collection

As the interview was expected to be quite long, the session was audio-recorded using a smartphone, while the interviewer was focused on obtaining as much information as possible from the interviewee. Recording the interview also allowed the interviewer to take notes regarding non-verbal communication that could complete some answers. As in the focus group, all participants were informed of the recording at the beginning of the session and were asked for permission. The interviewee was also informed about the planned use of the recording, which would only be analysing his/her answers. Finally, they were informed their name would not appear in any report or subsequent work without their explicit permission.

In this case, the interview was pilot-tested. In fact, two pilots were run. The first one was a controlled pilot where the interviewee was a fellow researcher, who was then ineligible for the study. The aim of this pilot was to set up all the materials and try out the whole procedure, to ensure that we had considered every relevant aspect and nothing was left to improvisation.
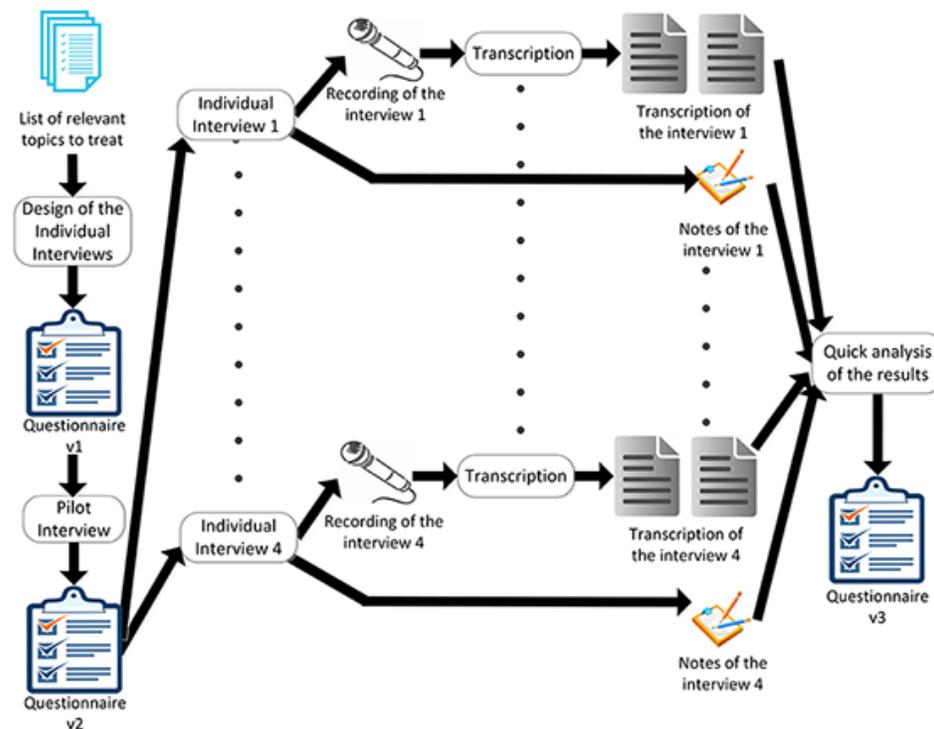

Figure 3: First round of interviews

Once the procedure was verified and refined according to the first pilot, we started running the

interviews with real participants. After conducting four interviews, the interviewer transcribed them in order to perform a first analysis evaluating how rich and complete the answers were, and if the questions were clear enough and covered the topic broadly. The analysis was made only by skimming the transcriptions of the interviews. This analysis, together with the experience of the interviewer, gave rise to reformulation and reordering for some of the questions of the questionnaire that were ambiguous or hard to understand for the interviewees. Nonetheless, no questions were added or deleted, keeping exactly the same content and semantic (see Figure 3).
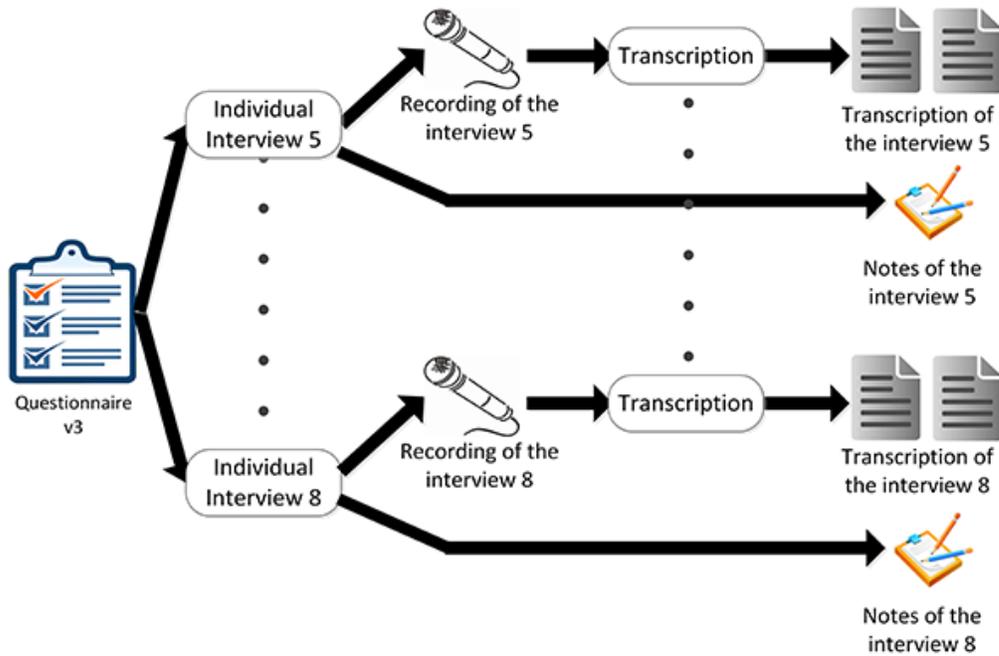

Figure 4: Second round of interviews

After this, four more researchers were interviewed with the updated questionnaire (see Figure 4). In both cases, each interview lasted around one hour, even if they were not time-limited and each interviewee had as much time as desired to express his/her ideas.

Apart from this, no further feedback was obtained from the participants —neither from further interviews, revision of the interviews' transcriptions, nor comments regarding the findings.

Data analysis

Our aim in this research is to provide a framework allowing the formalisation of the information-seeking process performed while researching in the computer science field. Amongst the available tools used to analyse qualitative studies, grounded theory (Glaser and Strauss, 2009) was used as it allowed us to start the analysis from scratch. Grounded theory is an inductive approach allowing concepts to emerge from freely, without a starting point of pre-defined concepts or hypothesis. With grounded theory a theory may emerge from the data, but it will not be proved (Corbin and Strauss, 2008).

The analysis was performed through a systematic and iterative approach based on the assignment of tags to parts of free text (Saldaña, 2012). These labels, named codes, are grouped according to the concept they deal with, which supports the formation of a theory. This codification and categorisation is performed iteratively in order to refine the analysis, typically following these four stages:

1. Open or emerging coding system

Text is coded from scratch, without using any kind of theoretical framework or published code list. One code is assigned to each unique phenomenon.

2. Development of concepts
   Similar codes are grouped under concepts.
3. Grouping concepts into categories
   Similar concepts are grouped into categories. Each category has a detailed description.
4. Formation of a theory
   Identification of connections or co-relations between concepts and categories in order to infer or predict a theory.

During recent decades, many researchers have proposed models of the information-seeking process. While some of them are rather general (Kuhlthau, 1991; Wilson, 1981; 1997; 1999), many of them only consider the information-seeking process in specific fields like social sciences (Ellis, 1989; Meho and Tibbo, 2003), physics (Cox, 1991), chemistry (Hall, 1991), engineering (Hertzum, 2000) or advocacy (Makri, Blandford, and Cox, 2008). Even if many of them also use grounded theory as the qualitative tool to formalise a theory (Ellis, 1993), none of them specifies which coding system they have used. In addition, to the best of the authors' knowledge, no models have been proposed for the information-seeking process in computer science, and there is no published coding system that could be useful as a base upon which to create an adapted one. Therefore we needed to develop our own coding system in order to formalise a model for the information-seeking process in computer science, and we consider that the resulting coding system could be of value for researchers who would like to perform a qualitative study related to this topic.

As of the writing of this article, we have completed the first three stages in the grounded theory method, and as a result a hierarchical list of codes grouped into categories of concepts has been generated. In the near future our intention is to further formulate a theory for the information-seeking process performed by researchers in computer science.

Codification process

Coders

Qualitative studies are inherently subjective because of their nature. There is no scientific and straightforward method to analyse text or speech, and then results, inevitably, depend to a greater or lesser extent on the researcher who carries out the codification.

Being aware of this limitation, we intended to provide results as objective as possible by minimizing individual subjectivity while coding as much as possible. To do this, we decided to code the interviews' transcript in parallel with three different coders. One of the coders was the same one who carried out the interviews and transcribed them, and may have had a higher risk for bias. To compensate for this, the other two coders did not participate in the qualitative study until all the interviews had been conducted and transcribed.

The two coders in charge of the validation of the coding system were recruited because of their closeness to the topic and because of their experience in conducting qualitative studies. Both coders have a similar profile: senior university professors and researchers in a computer science school, whose field of research is human-computer interaction, both around 40 years old. To avoid possible differences in interpretation of the interviews because of the gender, one of the coders was male and the other one was female.
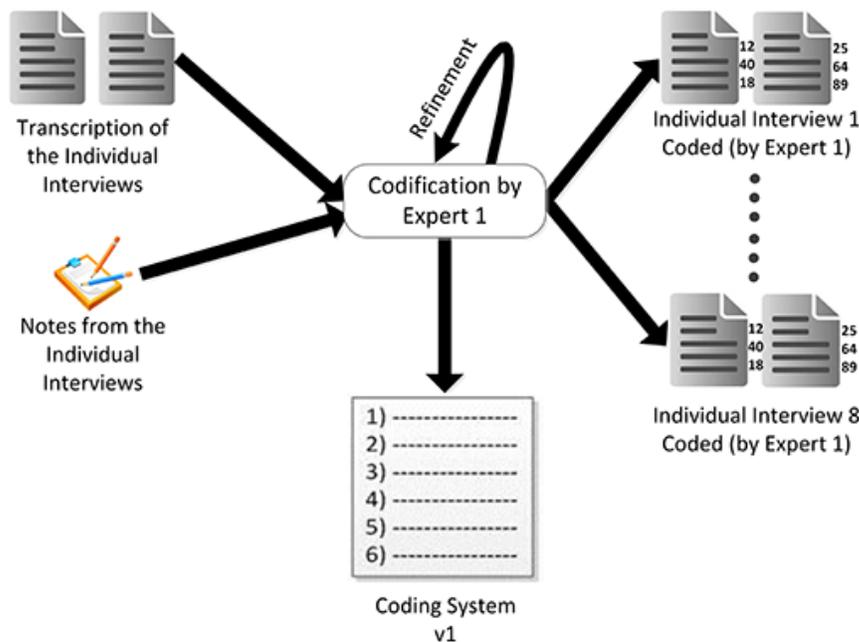
Iterative process

Figure 5: First tentative coding system

The codification process consisted of several phases. To begin with, the first coder developed a preliminary version of the coding system tagging the interviews' transcriptions totally from scratch (see Figure 5). The assignment of labels was performed using ATLAS.ti, a software tool for analysing data from qualitative studies. As we chose an interpretive approach to analyse the qualitative data, we needed codes to emerge from the text. Therefore, the coder labelled each sentence – or group of closely related sentences – with keywords that reflected the concepts or ideas emanating from the text. Initially, many of the codes that emerged were ad-hoc, redundant or hard to understand. An example is the code *'DIFFICULTIES – Find interesting articles or authors that do not have too much impact or are not known experts in the field' that was used to label the portion of text 'It is difficult to find somebody who is not very well known in the field or does not have published any article with a high impact, but is working in the same topic as I am'.* It seemed obvious that both the code and the text were almost the same, and then the code would be hardly reusable in other cases, offering little scope for re-use.

As the coder encoded the interviews, he acquired some perspective that allowed him to refine the initial codes and where they were used. Each iteration resulted in a new refinement to the coding system, where new codes were added and former codes were deleted or reformulated. As a result, when the coder concluded his work, the coding system contained 189 codes.
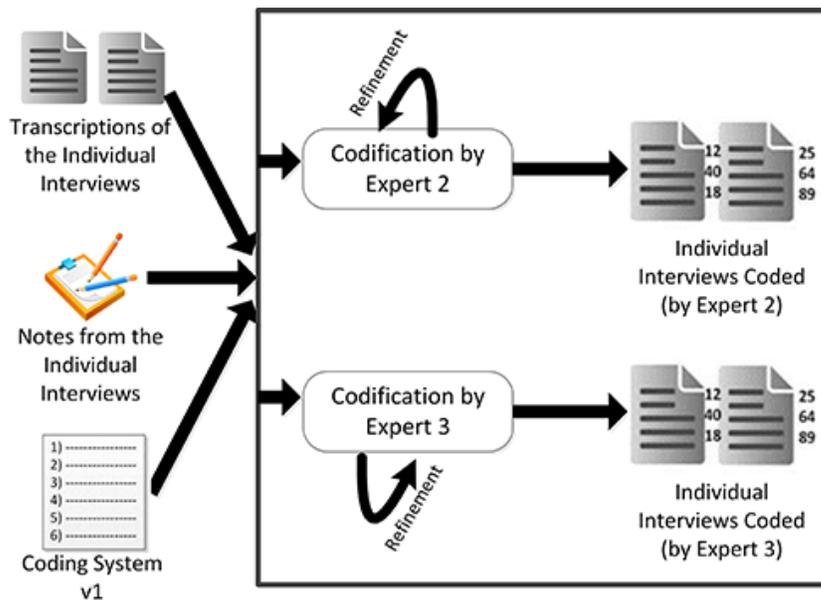
Figure 6: Second coding phase

In the second phase, the other two coders were given the transcriptions of the interviews alongside with the coding system, and were asked to code them individually, that is without contact between them (see Figure 6). No guidelines or rules were given to these coders to avoid influencing them. The only direction they were given was to identify bad codes – ambiguous, incomprehensible, misspelled – or relevant parts of text where no code was suitable to express the underlying idea. Furthermore, in order to avoid any external influence or preconceived ideas, transcriptions were anonymised. Coders were asked to revise and refine their codification until they thought it was acceptably accurate and comprehensive.
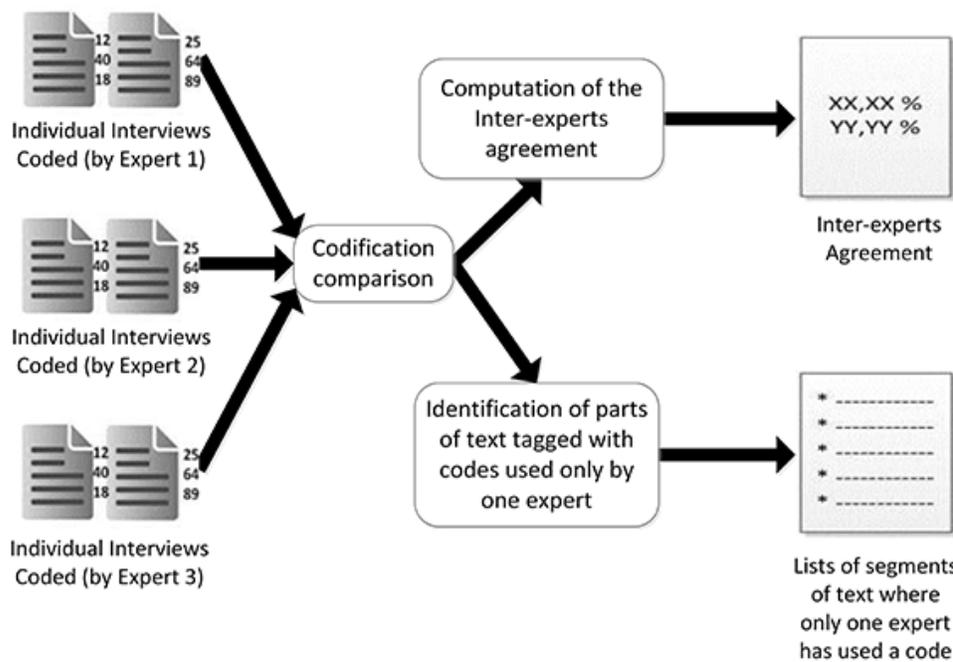
Figure 7: First refinement iteration: comparison of the coders' codification

Once all the interviews were coded by all the coders, the rate of agreement was computed to determine the validity of the coding system (see Figure 7). The computation was made taking into account how many codes had been used to tag the same portion of text. In this phase, the mean agreement rate of the three coders was 18.88%. In general terms, this rate may seem very

low, but considering that the codification is made for text without any kind of structure nor delimitation, that three different researchers have coded individually this text, that the codification was subjected to a high degree of subjectivity, and that a non-verified preliminary version of the coding system was used, the yardstick changes. For those reasons, we decided to adopt as the agreement rate the proportion of times that at least two of three coders have used the same codes to tag the same chunk of text. We considered that, if at least two coders decided individually to use the same code, this meant that the code was sufficiently descriptive and meaningful, and then had to be included in the coding system. Therefore in this second iteration the mean rate of agreement of at least two of the three coders was 40.87%, which is significantly higher than in the first iteration.
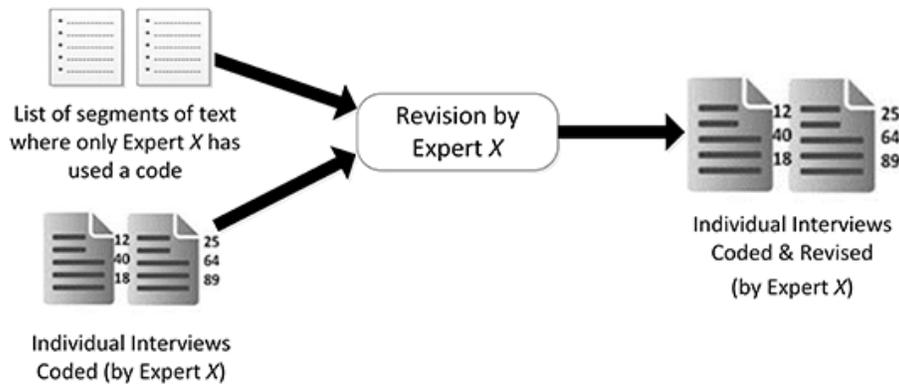

Figure 8: Second refinement iteration: individual refinement

A third iteration was planned from the beginning, as we expected that, this being their first approach to the codes, the two coders validating the coding system would need some time and practice to become familiar with them. Thus, in this iteration, the starting point was the three sets of coded interviews, each one coded by each of the coders. The main objective in this phase was to identify if differences in codification were due to errors while coding, misunderstandings, or because of differences in interpretation. Each coder received his/her set of coded interviews together with the list of codes that had been used only by him/her to tag a specific portion of text. Each coder then had to evaluate, again individually, why they were the only ones using these codes. As a result, coders could reconsider their decision and use another code they consider more appropriate, or on the contrary they could reaffirm their decision and maintain the original code (see Figure 8). After this revision, the mean agreement rate was 28.53% for all three coders when all the codifications were considered, while 57.53% of the codification was the same for at least two of the three coders.
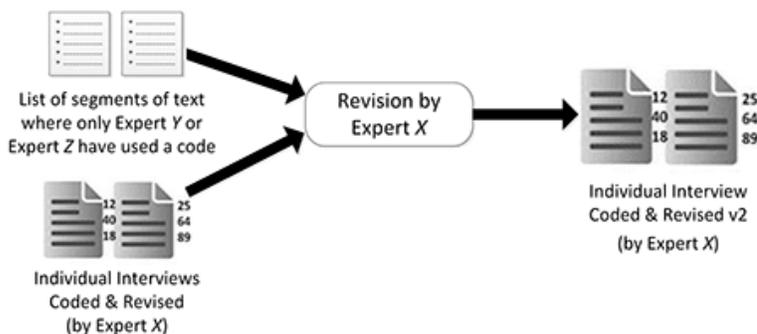

Figure 9: Third refinement iteration: refinement by the other experts

The next step consisted of identifying which codes were still used by only one coder in a specific chunk of text, but in this case we sent the codes and the text where they were used to the other two coders (see Figure 9). At this point, if at least one of these two coders agreed with the use of

the code, it was validated and added to the coding system. Conversely, if both coders were against using the code in that portion of text, the code was discarded. This validation phase resulted in a 31.18% of agreement between all the coders, and 66.43% between at least two of the three coders.
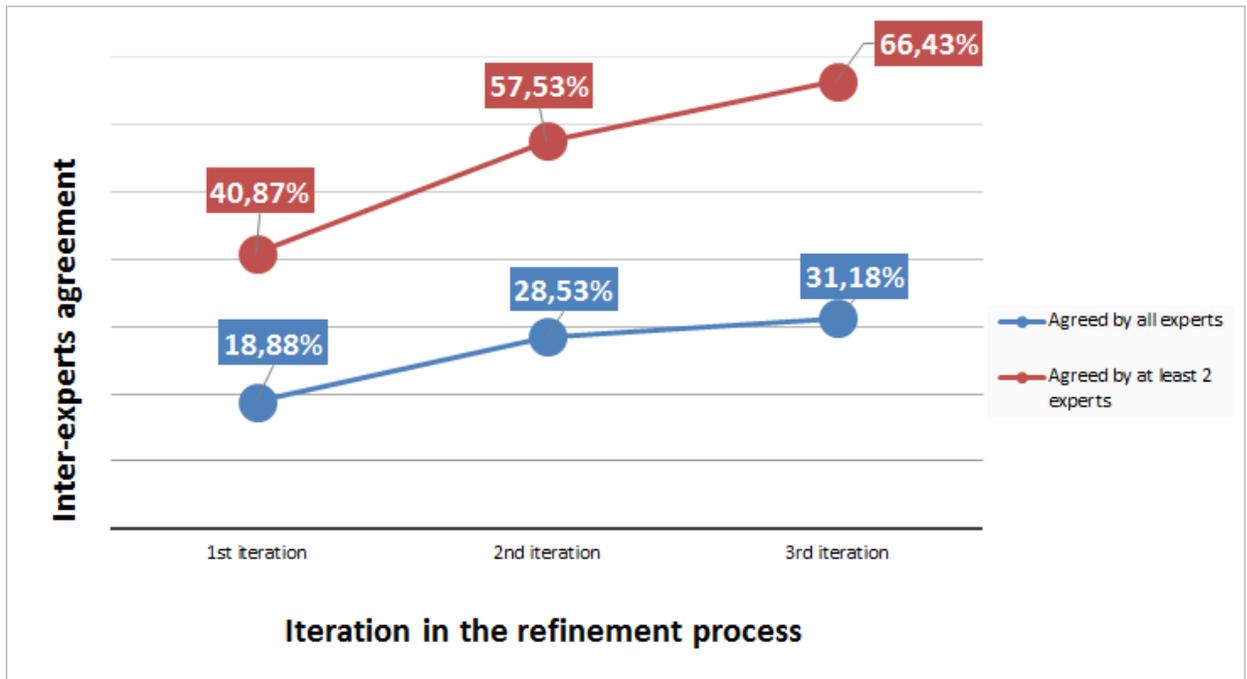
Figure 10: Evolution of the inter-coders agreement during the refinement process

Figure 10 shows the evolution of both inter-coders agreement rates – rate of codes agreed by all the coders, and rate of codes agreed by at least two of the three coders – during the refinement process.
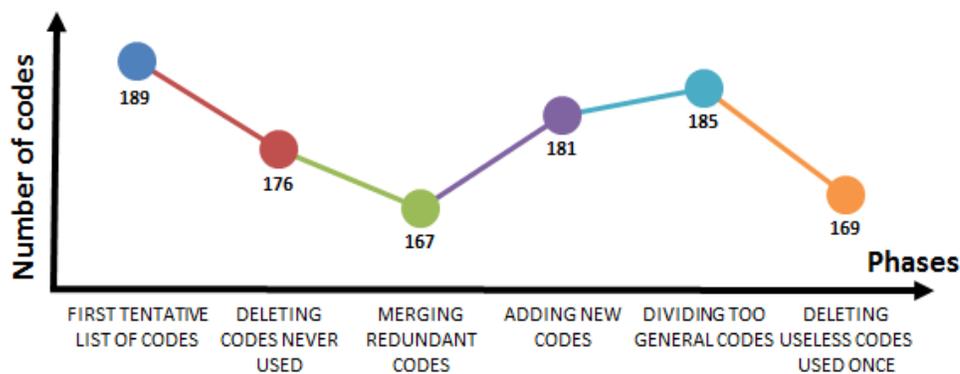
Figure 11: Evolution of the coding system

Action Performed	Initial Code(s)	Resuting Code(s)
DELETING CODES NEVER USED	WISHES: Integrate all the available information.	-
MERGING REDUNDANT CODES	WORKSPACE: Relevant articles are clearly differentiated. WORKSPACE: Relevant articles are kept at hand.	WORKSPACE: Relevant articles for the current task are kept at hand.
ADDING		SEARCH - SOURCES OF

NEW CODES	-	INFORMATION: Local Repository.
DIVIDING TOO GENERAL CODES	WISHES: Suggestions should be displayed only on demand and in a non-intrusive way.	USER WISHES - VISUALIZATION: To display suggestions only on demand.
		USER WISHES - VISUALIZATION: To have a non-intrusive visualisation of the suggestions.
DELETING USELESS CODES USED ONCE	INFORMATION-SEEKING PROCESS: It is the starting point to write an article.	-

Table 1: Examples of actions performed over the coding system in the last iteration.

Finally, a last iteration was performed to filter the codes that had little or no use by the coders. Additionally, new codes had to be added to the first version of the coding system in order to label the portion of text, identified by the two coders in charge of the validation of the coding system, where no previous code was good enough to label the underlying idea. This phase was carried out over several group meetings, where decisions were discussed and agreed to by all three coders. The codes that had not been used at all in the final codification of the interviews were considered redundant, erroneous or unneeded, so they were deleted from the coding system. Altogether, 13 codes were deleted, with a total of 176 codes left in the coding system. After this step, 17 codes, which had been used at least one time, were considered redundant and merged into 8 new codes. This merge decreased the number of codes to 167. After this, all three coders discussed proposed new codes not present in the first coding system. The result was that 14 codes were added to the coding system, increasing the total number of codes to 181. Some codes were identified as being too abstract, and we decided to divide them into more specific codes: two of the former codes were divided each into two new codes, while a third former code gave way to three new codes. From the remaining 185 codes, 31 of them had been used to code only one text segment among all the interviews. Between these codes, the research team determined that 16 of them were unnecessarily used – for example because they were too specific, or because they did not express a relevant idea, and then had to be deleted, while others were representative of very specific and quirky, but also relevant concepts that had to be represented in the coding system. The evolution of the coding system in this last iteration has been depicted in the Figure 11, and Table 1 provides some examples of the actions performed over the codes.

Results

At the end, we had a coding system that included 169 codes. In order to facilitate the understanding and further use of these codes, for each of them we have written a short description that clarifies its meaning. Additionally, we provide an example of use, extracted from the interviews, for each of the proposed codes, so as to foster application, indicating from which interviewee and at which line of the transcription the text comes from. The full list of codes with their description and examples of use are detailed in Appendix A.
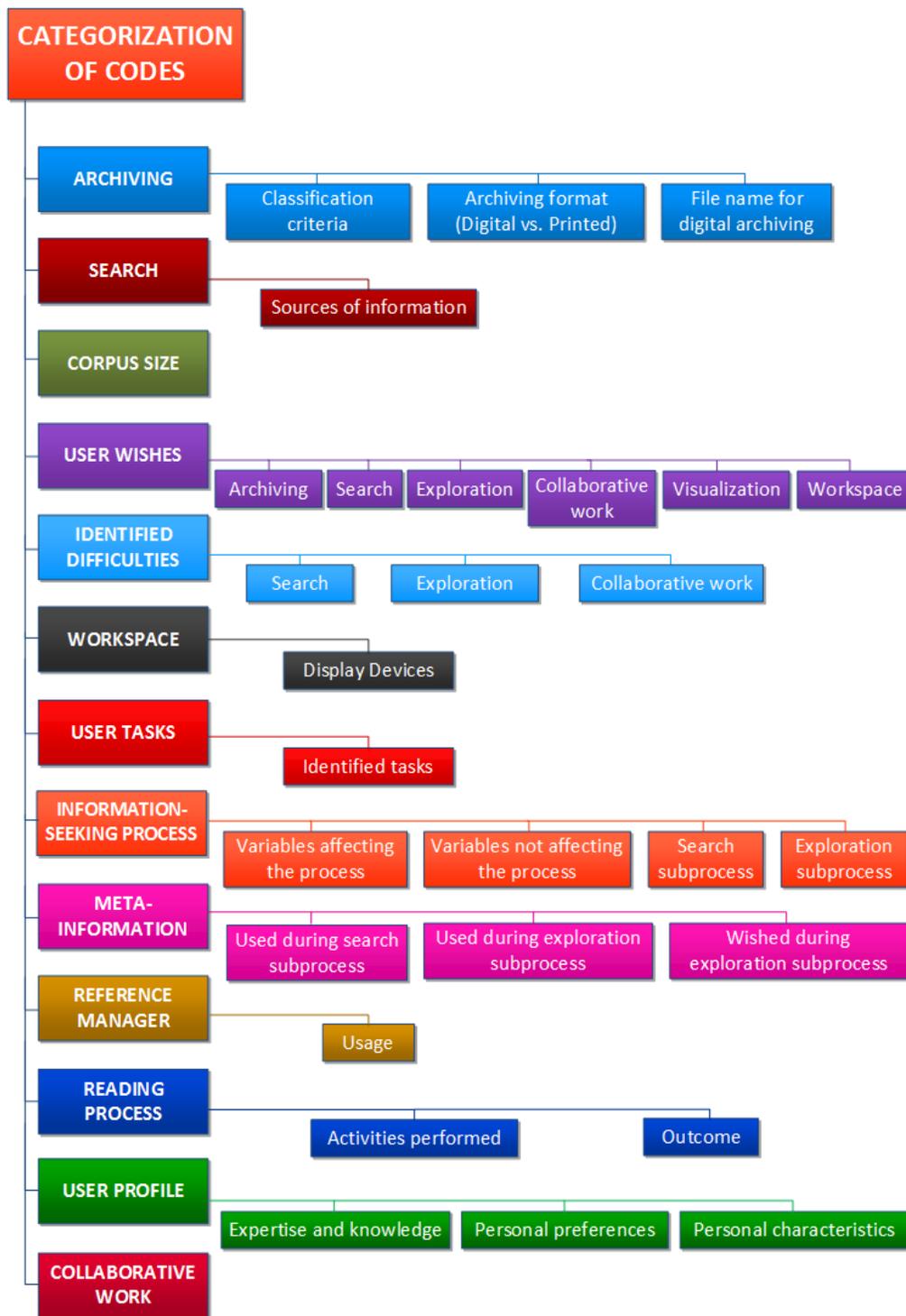
Figure 12: Hierarchical organization of the codes

Additionally, we have hierarchically categorised all the codes to allow potential users of the coding system to locate the desired code(s) more easily (see Figure 12). To elaborate these categories, we have taken into consideration the initial aspects identified in the focus group, but also the codes themselves and their use while coding the interviews. Some of these categories were in turn divided into more specific subcategories, which allowed better grouping of the codes based on the specific aspect they dealt with. As an example, the category Archiving contains three subcategories: *Classification criteria*, *Archiving format (Digital vs. Printed)* and *File name for digital archiving*.

Conclusions and future works

Information-seeking is a complex process to define because it consists of a variety of tasks, as well as the large amount of information that is available and accessible. Designing a system to support such process is a challenge mainly because it needs to provide a high degree of interaction with the users. These users have different mental models, different skills, different habits and different goals. In brief, users introduce into the process a high degree of variability. A system that does not consider this variability will oblige the user to adapt his/her methods, goals and personal preferences to the system, instead supporting the user by customizing the system to the user's needs.

Our purpose is to identify the commonalities and specificities, both at the process level and at the user level, to design a system that can adapt its functionality, visualisation and interaction based on the task at hand, the user and the context of use. With this aim, we have run a set of qualitative studies to obtain real information from real users.

In this paper we present the first result obtained after analysing the data, which is a coding or text tagging system for researchers who qualitatively study information seeking processes in computer science.

In the immediate future, we will continue to work on the generation of a set of conceptual models, including those of the information seeking process and the researcher, among others.

As every qualitative study has its own characteristics, restrictions and purposes, it would be impossible to cover every single situation, so we will not claim our coding system is universally applicable. Nevertheless, it can be useful as a reference tool to conduct similar or related qualitative studies. Indeed we plan to evaluate the degree of flexibility of the proposed coding system when applied to new qualitative studies of the information-seeking process in other fields of research.

Acknowledgements

About the authors

Cristian Moral is a Ph.D. Candidate and Associate Professor in the Escuela Técnica Superior de Ingenieros Informáticos (ETSIINF) at Universidad Politécnica de Madrid (UPM), Spain. He received his Master of Science in Computer Science both from Universidad Politécnica de Madrid and from Politecnico di Torino (Italy). His current research interests are information retrieval, visualisation and manipulation through adaptive virtual environments, and human-computer interaction. He can be contacted at: cmoral@fi.upm.es.

Angelica de Antonio has been faculty member in the Escuela Técnica Superior de Ingenieros Informáticos (ETSIINF) at the Universidad Politécnica de Madrid (UPM) since 1990. She received her PhD in Computer Science in 1994. Her current research interests focus on virtual and augmented reality, adaptive systems and human-computer interaction. She can be contacted at: angelica@fi.upm.es.

Xavier Ferre has been faculty member in the Escuela Técnica Superior de Ingenieros Informáticos (ETSIINF) at the Universidad Politécnica de Madrid (UPM) since 1999. He

received his PhD in Computer Science in 2005. His primary research interests are interaction design and user experience evaluation in mobile applications, user experience in e-health mobile applications, and integration of human-computer interaction practices into software engineering. He can be contacted at: xavier.ferre@upm.es.

Graciela Lara is a Ph.D. Candidate in the Escuela Técnica Superior de Ingenieros Informáticos (ETSIINF) at Universidad Politécnica de Madrid (UPM), Spain. She received a PhD in Teaching Methodology in 2012 and is an associate professor at the Computer Science Department at the CUCEI of the Universidad de Guadalajara, Mexico. Her current research interests are virtual reality, mainly on its application for training, 3D object modelling, and spatial mental models. She can be contacted at: graciela.lara@red.cucei.udg.mx.

Bernard, H. (1988). Research methods in cultural anthropology. Newbury Park, California: SAGE Publications, Inc.

Corbin, J. & Strauss, A. (2008). Basics of qualitative research (3rd. ed.). Los Angeles: SAGE Publications, Inc.

Cox, D. (1991). An investigation into the information seeking behaviour and needs. Sheffield: University of Sheffield, Department of Information Studies. Retrieved 28 October 2015 from http://bit.ly/1MmqCA1 (Archived by WebCite® at http://www.webcitation.org/6ccaFT7AU)

Crabtree, B. F. & Miller, W. L. (1999). Doing qualitative research (2nd. ed.). London: SAGE Publications, Inc.

Ellis, D. A. (1989). A behavioural approach to information retrieval system design. *Journal of Documentation*, **45**(3), 171-212.

Ellis, D. A. (1993). Modeling the information-seeking patterns of academic researchers: a grounded theory approach. *Library Quarterly*, **63**(4), 469-486.

Glaser, B. G. & Strauss, A. L. (2009). The discovery of grounded theory: strategies for qualitative research. New Brunswick, NJ: Transaction Publishers.

Hall, A. A. (1991). A behavioural model of the information seeking behaviour of the academic chemists at the University of Sheffield. Sheffield: University of Sheffield, Department of Information Studies.

Hertzum, M. (2000). The information-seeking practices of engineers: searching for documents as well as for people. *Information Processing & Management*, **36**(5), 761-778. Retrieved 28 October 2015 from http://bit.ly/1kYptIX (Archived by WebCite® at http://www.webcitation.org/6ccaegBFf)

Kitzinger, J. (1994). The methodology of focus groups: the importance of interaction between research participants. *Sociology of Health & Illness*, **16**(1), 103-121. Retrieved 28 October 2015 from http://bit.ly/1WlDX6H (Archived by WebCite® at http://www.webcitation.org/6ccau4dQ9)

Kitzinger, J. (1995). Qualitative research: introducing focus groups. *BMJ*, **311**(7000), 299-302.

Krueger, R. A. & Casey, M. (2009). Focus groups: a practical guide for applied research (4th. ed.). Los Angeles: SAGE Publications, Inc.

Kuhlthau, C. C. (1991). Inside the search process: information seeking from the user's perspective. *Journal of the American Society for Information Science*, **42**(5), 361-371. Retrieved 28 October 2015 from https://comminfo.rutgers.edu/~kuhlthau/docs/InsidetheSearchProcess.pdf (Archived by WebCite® at http://www.webcitation.org/6ccbGEuN6)

Makri, S., Blandford, A. & Cox, A. L. (2008). Investigating the information-seeking behaviour of academic lawyers: from Ellis's model to design. *Information Processing & Management*, **44**(2), 613-634. Retrieved 28 October 2015 from http://bit.ly/1KKrptp (Archived by WebCite® at http://www.webcitation.org/6ccbS14ph)

Meho, L. I. & Tibbo, H. R. (2003). Modeling the information-seeking behavior of social

scientists: Ellis's study revisited. *Journal of the American Society for Information Science and Technology*, **54**(6), 570-587.

Miles, M. B. & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook. Los Angeles: Sage Publications, Inc.

Powell, R. A. & Single, H. M. (1996). Focus groups. *International Journal for Quality in Health Care*, **8**(5), 499-504. Retrieved 28 October 2015 from http://intqhc.oxfordjournals.org/content/intqhc/8/5/499.full.pdf (Archived by WebCite® at http://www.webcitation.org/6ccc7rNp7)

Saldaña, J. (2012). The coding manual for qualitative researchers (2nd. ed.). London: SAGE Publications, Inc.

Tong, A., Sainsbury, P. & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, **19**(6), 349-357. Retrieved 28 October 2015 from http://intqhc.oxfordjournals.org/content/intqhc/19/6/349.full.pdf (Archived by WebCite® at http://www.webcitation.org/6cccHCenS)

Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation*, **37**(1), 3-15. Retrieved 28 October 2015 from http://www.informationr.net/tdw/publ/papers/1981infoneeds.html (Archived by WebCite® at http://www.webcitation.org/6cccQIt2c)

Wilson, T. D. (1997). Information behaviour: an interdisciplinary perspective. *Information Processing & Management*, **33**(4), 551-572.

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation*, **55**(3), 249-270. Retrieved 28 October 2015 from http://www.informationr.net/tdw/publ/papers/1999JDoc.html (Archived by WebCite® at http://www.webcitation.org/6cccrTxF9)

Find other papers on this subject

Check for citations, using Google Scholar

Appendix A: Categorised coding system

For each code (underlined font) we provide its definition (normal font) and an example. Each example is formed by a couple of question (bold and italic font) and answer (italic font) extracted from one of the eight interviews performed. In most cases, the full answer to the question has not been included, but only the sentence(s) that were tagged with the code in question. The question is provided to set the answer in context.

ARCHIVING
1. Article for possible future use
The interviewee refers to an article that may be relevant in the future, but not at the moment.
If you find an interesting article that is related to your field of study but that it is not

> *useful for what you are looking for at that moment, what do you do with it?*
> *Usually I save it in my computer, inside a folder named "Secondary".*

2. Hierarchical organization

> The interviewee mentions that he/she uses a hierarchical scheme to classify articles.
> *How do you store he articles once you have selected them?*
> *I have a folder for each project. Within each project's folder, I organize the papers according to their topics, going from the most general topic to the most specific one.*

Classification Criteria

3. Specific title

> The interviewee mentions that he/she names the folder with a specific user-defined title.
> *How do you store he articles once you have selected them?*
> *I have a folder called "Others" where I store the papers I don't know how to classify.*

4. Year of publication

> The interviewee mentions that he/she names the folder with the year of publication of the articles.
> *How do you store he articles once you have selected them?*
> *I keep the references sorted by publication year in a bibtex file in CiteULike.*

5. Author

> The interviewee mentions that he/she names the folder with the name of the author(s).
> *How do you store he articles once you have selected them?*
> *At first I had a folder named "Interesting papers", but recently I have added some subfolders, according to the publication venue or the author of the articles, because I had too many articles and it was starting to become unmanageable.*

6. Publication venue

> The interviewee mentions that he/she names the folder with the name of the publication venue (journal, conference, workshop...).
> *How do you store he articles once you have selected them?*
> *At first I had a folder named "Interesting papers", but recently I have added some subfolders, according to the publication venue or the author of the articles, because I had too many articles and it was starting to become unmanageable.*

7. Intended Purpose

> The interviewee mentions that he/she names the folder according to the aim for which he/she plans to use the articles.
> *How do you store he articles once you have selected them?*
> *My main sorting criterion is the intended purpose for the article, since I have a folder for every article I write.*

8. Subject

> The interviewee mentions that he/she names the folder with the articles subject.
> *How do you store he articles once you have selected them?*
> *I sort the papers according to their intended purpose: teaching, research... As a second sorting criterion I use the article subject, going from the most generic articles to the most specific ones.*

Archiving format (Digital vs. Printed)

9. Both digital and printed

The interviewee mentions saving articles in both formats: digital and printed.
> *How do you store he articles once you have selected them?*
> *I store it in digital format, but when I need to read it, I print it. I only read in paper.*

10. Only digital

> The interviewee mentions saving articles only in digital format.
> *How do you store he articles once you have selected them?*
> *I store it in digital format.*

11. Only paper

> The interviewee mentions saving articles only as printed documents.
> *If you find an interesting article that is related to your field of study but that it is not useful for what you are looking for at that moment, what do you do with it?*
> *The ones I have found while looking for a different thing, I usually print them to read them at any other time.*

12. Only relevant articles are printed

> The interviewee mentions printing articles only if he/she considers them relevant.
> *How do you store he articles once you have selected them?*
> *I save it in digital format, and if it is important I also print it.*

File name for digital archiving

13. Original file name

The interviewee mentions that he/she keeps the file name provided by the source where the article was obtained.
> *If you find an interesting article that is related to your field of study but that it is not useful for what you are looking for at that moment, what do you do with it?*

I save it directly with the title it has when I downloaded it.

14. <u>Publication year of the article</u>
The interviewee mentions that he/she includes the year of publication of the article in the file name.

> *How do you store he articles once you have selected them?*
> *I usually rename it with the author and the year of publication.*

15. <u>Author of the article</u>
The interviewee mentions that he/she includes the article author(s) in the file name.

> *How do you store he articles once you have selected them?*
> *I usually rename it with the author and the year of publication.*

16. <u>Title of the article</u>
The interviewee mentions that he/she includes the title of the article in the file name.

> *How do you store he articles once you have selected them?*
> *I save it using the format of a bibliographic reference as file name, in order to be able to see at first sight the information that allows to quickly identify the article: author (using 'et al.' if there are several authors), year and title.*

17. <u>Publication venue of the article</u>
The interviewee mentions that he/she includes the publication venue of the article in the file name.

> *How do you store he articles once you have selected them?*
> *I save it with the name of the author I know, or the name of the main author if I don't know any of the authors, and a few keywords that summarise the paper. In many cases, I also add the publication venue.*

18. <u>Reference manager</u>
The interviewee mentions that he/she includes the name of the tool used to manage the article in the file name.

> *How do you store he articles once you have selected them?*
> *I save it in my computer, using as filename the title of the paper and the initials of the reference manager - if I have uploaded the paper to it - so that I know I have already processed it.*

19. <u>Self-defined keywords</u>
The interviewee mentions that he/she includes self-defined keywords in the file name.

> *How do you store he articles once you have selected them?*
> *I save it with the name of the author I know, or the name of the main author if I don't know any of the authors, and a few keywords that summarise the paper. In many cases, I also add the publication venue.*

SEARCH
Sources of information
20. <u>Search engines / repositories are the starting point of the search process</u>
The interviewee mentions that his/her search process starts with search engines and/or repositories, in the sense that they are the basis of the process.

> *How do you perform the information-seeking process?*
> *I first launch searches in general search engines to cover as many results as possible. Once I have a certain idea of which terms I should use in a query, I use scientific search engines like CiteSeer or the Web Of Knowledge.*

21. <u>Search engines / repositories are not the basis of the search process</u>
The interviewee mentions using search engines and/or repositories as a means to obtain complete articles, but not as the basis of the search process.

> *How do you perform the information-seeking process?*
> *Now that I am an expert in my topic, the procedure has changed because I don't use generic repositories anymore, but I consult directly the relevant journal in my area, and I read all the latest issues.*

22. <u>Expert</u>
The interviewee mentions that he/she requires help from an expert in the subject he/she is researching.

> *How do you perform the information-seeking process?*
> *Before anything else, I speak with my tutor to debate which topics or range of topics we want to treat for a possible publication, and from there I obtain a list of topics, journals, conferences... to treat and consult.*

23. <u>Local repository</u>
The interviewee mentions that he/she carries out searches in his/her repository of articles.

> *How do you perform the information-seeking process?*
> *If what I am looking for sounds familiar to me or if I want to find a specific paper I have already read, but I don't remember which article it is, I perform a search in my document of summaries. This is useful only if I have been methodical and have saved a summary of the article in question in it.*

24. Use of different kinds of sources

The interviewee mentions that the type of sources of information he/she uses varies (e.g. general search engine, specific conference, colleagues...) during the search process

> *How do you perform the information-seeking process?*
> *It depends a little bit on the purpose of the search: if I don't know the topic, I start the search using general search engines, especially to identify the terminology that experts usually use; if I need a more specific and theoretical information, I turn to a book; if I need a reference to use as a citation while I am writing a paper, I directly use a scientific search engine.*

CORPUS SIZE

25. The exploration process varies according to the corpus size

The interviewee mentions that the exploration process of the results returned by a search depends on the amount of results obtained.

> *Does the exploration of information vary according to the number of search results? If so, how?*
> *If the number of papers to explore is manageable, I can afford to be exhaustive and read all of them one by one. If there are too many, it is necessary to refine the search improving the query and performing many levels of filtering, for example taking into account the title, the authors, the abstract, the introduction or even, in some cases, the whole paper. This way, I discard the papers that I believe are not interesting.*

26. The exploration process does not vary according to the corpus size

The interviewee mentions that the exploration process of the results returned by a search is always the same, regardless of the amount of results obtained.

> *Does the exploration of information vary according to the number of search results? If so, how?*
> *No, I always do the same: I look for the papers and then I filter them based on their publication venue and their authors. After this, if I like any of them, I read the abstract to decide if it can be useful or not.*

27. Detailed analysis of the corpus when its size is manageable

The interviewee mentions that, if the size of the set of results returned by a search is small enough to be manageable, then he/she analyses each of them more in detail.

> *Does the exploration of information vary according to the number of search results? If so, how?*
> *Of course it varies. The more the corpus is small, the more I can afford reading more in detail. In this case, it is worthy to look at all of them, one by one, more in detail and read the title, the abstract and even the introduction.*

28. Filtering is applied to the corpus when its size is not manageable

The interviewee mentions that, if the size of the set of results is too big to be manageable, then he/she filters these results until the size is manageable.

> *Does the exploration of information vary according to the number of search results? If so, how?*
> *If the amount of results is too high, first of all I filter them based on their topics or on other criteria until the set has a manageable size.*

USER WISHES
Archiving

29. To add user-defined tags to articles

The interviewee would like to be able to add his/her own defined tags to articles.

> *Do you miss something during the information-seeking process? If so, what?*
> *I would like to be able to organize the papers using labels, like in Gmail, because the same paper can be related to several topics or because sometimes it could be useful to tag a paper with its topics and the conference where it has been published, for example.*

30. The system should help to classify articles

The interviewee would like the system to help him/her to classify articles.

> *Do you miss something during the information-seeking process? If so, what?*
> *I would like to be able to classify more easily the papers that I consider useful.*

31. The system should classify articles automatically

The interviewee would like the system to classify articles automatically.

> *Do you miss something during the information-seeking process? If so, what?*
> *The system could organize automatically the papers I have downloaded and then facilitate me to retrieve them back more easily.*

32. To save automatically an article (or a reference to it) in a remote system

The interviewee would like to be able to directly save an article found during the information-seeking process in a remote system (reference manager, cloud storage...).

> *Do you miss something during the information-seeking process? If so, what?*
> *I would like that the tools were more integrated in terms of management of the*

documents and references, in order to avoid making me do some tasks manually, for example downloading the papers from a scientific repository and then uploading them manually to the reference manager.

Search

33. <u>To perform high-level searches</u>

The interviewee would like to be able to use more general terms for constructing the search query (topics, concepts...). This means that the system should be able to return related articles, even if the terms of the query are not present in these articles.

Do you miss something during the information-seeking process? If so, what?
The papers could be better organized using their topics as the classification criterion, so that I could use topics in the query instead of keywords.

34. <u>The system should help to search in a local repository</u>

The interviewee would like to be able to have support to perform searches in his/her local collection of documents.

Do you miss something during the information-seeking process? If so, what?
To be able to manage more easily my local bibliography and to be able to use it as a source of information to perform searches, for example using some filters.

35. <u>The system should help to explore search results</u>

The interviewee would like that the system offers support for the process of exploration of search results.

Do you miss something during the information-seeking process? If so, what?
I would like the system to act like an expert who would give me some guidelines to explore the search results, especially if the topics that I am investigating are unknown for me, for example offering me some initial informative indications.

36. <u>The system should offer suggestions of interest</u>

The interviewee would like that the system offers suggestions relevant for his/her research (alternative terms for a query, related authors, relevant venues, reference articles...).

Do you miss something during the information-seeking process? If so, what?
The system could suggest me articles based on my interests and on what other researchers with similar interests have already searched. It's something I usually use when I buy books, and then I think it would be interesting.

Exploration

37. <u>The system should keep a history of the user exploration</u>

The interviewee would like that the system creates and updates a user history based on which articles have been downloaded, have already appeared / not appeared in previous search results...

Do you miss something during the information-seeking process? If so, what?
The search engine could save my exploration history in order to indicate me, or even automatically filter for me, which papers I have already seen, which ones I have already discarded or even which ones are new and have never appeared to me before.

Collaborative work

38. <u>To personalise the interaction and the visualisation without affecting the rest of the team</u>

The interviewee would like that visualisation and interaction can be customised individually, without affecting the rest of the team members.

Do you miss something during the information-seeking process? If so, what?
It would be nice if each member of the team could customise how he/she wants to visualise and manage the information, for example naming a file with his/her own terms or using his/her own classification, but always without affecting the rest of the team.

39. <u>To get more support for communication between team members</u>

The interviewee would like the system to provide more support to facilitate and enhance the communication between team members.

Do you miss something during the information-seeking process? If so, what?
I would like the system would provide more support to the collaborative exploration of the documents, so that I could know what the other team members are working on and which comments they have made about the papers they have already read, or to be able to talk with them using a paper as a common base.

40. <u>To have access to the exploration history of other team members</u>

The interviewee would like to know which articles other team members have already downloaded, analysed, rejected...

Do you miss something during the information-seeking process? If so, what?
I wonder what my colleagues, friends or other researchers have read and which are their opinions about what they have read.

41. <u>To have access to annotations added by other team members</u>

The interviewee would like to see the annotations (comments, schemes, tags...) added by other team members to an article.

Do you miss something during the information-seeking process? If so, what?
I wish it was easier to access the papers that other team members have already

downloaded, and I would like to be able to see the comments and notes they could have written about them.

42. <u>The system should offer suggestions of interest</u>
The interviewee would like to have access to the articles that other team members have saved.
> *Do you miss something during the information-seeking process? If so, what?*
> *I wish it was easier to access the papers that other team members have already downloaded, and I would like to be able to see the comments and notes they could have written about them.*

Visualisation

43. <u>To display meta-information only on demand</u>
The interviewee would like to decide when he/she wants to have meta-information displayed.
> *Do you miss something during the information-seeking process? If so, what?*
> *Mainly I would love to know what are the general concepts addressed in a paper, because this would spare me from having to read the title, the abstract, the conclusions... The concepts would suffice to determine if a paper is interesting or not. Now, it should always be on demand, so that each user could decide what to use at every moment, maintaining the control of the process.*

44. <u>To display suggestions only on demand</u>
The interviewee would like to decide when he/she wants to get suggestions from the system.
> *Do you miss something during the information-seeking process? If so, what?*
> *The system could suggest things, but in a non-invasive way, that is allowing me to maintain the control and the power to decide if I want or not to have a suggestion, and without requiring from me an extra work in order to discard a suggestion (for example a pop-up window).*

45. <u>To have a non-intrusive visualisation of the suggestions</u>
The interviewee would like the suggestions to be shown non-intrusively (i.e. without blocking the current window, without breaking the flow of the current task...).
> *Do you miss something during the information-seeking process? If so, what?*
> *I wish the system would offer me recommendations based on my history but only on demand and in a non-intrusive way, as I want to be the one who decides if a suggestion is valid. Besides, I should be able to do it whenever I wanted, not when the system decides it.*

46. <u>To customise the visualisation and interaction</u>
The interviewee would like to customise the visualisation and interaction.
> *Do you miss something during the information-seeking process? If so, what?*
> *I would like to be able to more explicitly see which are the relations between the articles, provided that I could dynamically define which type of relations I want to see.*

47. <u>To visualise the collection as a network</u>
The interviewee would like to see the documents as visual elements that are connected, forming a network where connections between documents are clearly recognisable.
> *Do you miss something during the information-seeking process? If so, what?*
> *I would like that the relations between the articles were shown graphically and on demand, especially the network of authors and the network of references.*

48. <u>To interact directly with the visual representation of documents</u>
The interviewee would like to interact with the documents of the collection by manipulating their visual representation.
> *Do you miss something during the information-seeking process? If so, what?*
> *I would like the results were displayed as a network. I imagine a network of spheres, and when selecting one of these spheres, its relations with other spheres would appear graphically in the display, letting me indicate which spheres I'm interested in and which ones I'd like to delete in order to clear the network.*

Workspace

49. <u>To have more than one workspace</u>
The interviewee would like to have more than one workspace at the same time to associate each of them to a different task or purpose. For example, one workspace for interesting articles and another for notes.
> *Do you miss something during the information-seeking process? If so, what?*
> *I would like to have several workspaces in order to separate the interesting papers from the rest of papers, or for other kind of tasks.*

WORKSPACE

50. <u>It varies according to the specific task at hand</u>
The interviewee mentions that his/her workspace changes depending on the specific task at hand.
> *How is your workspace set up during the information-seeking process?*
> *If the information-seeking process is prior to the writing of the article, I only use on screen, whereas if I do the search during the writing phase, I have in one screen the text*

processor and in the other one I perform precise and specific searches of information or reference papers.

51. It does not vary according to the specific task at hand

The interviewee mentions that his/her workspace is the same regardless of the specific task at hand.

How is your workspace set up during the information-seeking process?
I always use only one screen because I prefer to concentrate only on one thing at a time.

52. It is divided in several zones

The interviewee mentions that he/she divides the workspace into different zones.

How is your workspace set up during the information-seeking process?
I have two displays. One of them has the latex editor, while in the other are the folders with the papers related with the topic of the paper I'm writing. Besides, I have at hand the most relevant papers printed.

53. Relevant articles for the current task are kept at hand

The interviewee mentions that he/she keeps the articles that are important for the current task at hand.

How is your workspace set up during the information-seeking process?
On the desk I have the articles I am currently working on.

54. Personal notes are kept at hand

The interviewee mentions that he/she keeps the written notes about references at hand.

How is your workspace set up during the information-seeking process?
I have at hand the notes and schemes I have written down while reading the papers.

55. A part of the workspace is dedicated specifically to write notes

The interviewee mentions that he/she reserves a certain part of the workspace for writing notes, regardless of the current task.

How is your workspace set up during the information-seeking process?
I always have paper sheets at hand in order to write the ideas I can have during the information-seeking process.

Display Devices

56. Only one screen

The interviewee mentions that he/she uses just one computer screen.

How is your workspace set up during the information-seeking process?
I only use one display, but with several desktops. In fact, I think I couldn't use more than one screen at a time.

57. More than one screen

The interviewee mentions that he/she uses at least two computer screens.

How is your workspace set up during the information-seeking process?
I have two displays. In the main display, I have the text editor and the references I am going to use to write my paper, whereas in the other display I have other things I can need, like a web browser.

58. Tablet

The interviewee mentions that he/she uses a tablet.

How is your workspace set up during the information-seeking process?
I use my tablet to store and classify the papers, but usually I read them on paper.

USER TASKS

59. Different tasks identified

The interviewee mentions that there are different tasks during the research process that require searching & exploring information.

When do you need to seek for information?
First of all I need to find where I can publish my paper. Then, while I am writing my article, I need to search and explore again in order to obtain more specific information, for example the theoretical framework, to contrast an hypothesis, or to cite other authors.

Identified tasks

60. To update article's references

The interviewee considers that one of the possible tasks in the information-seeking process is to find new references that are relevant for his research, but that he may have not found before because they may have been published recently.

When do you need to seek for information?
Another moment is when I want to publish a paper that I have written some time before. In this case, I have to perform a quick search to update it and include potential new references in the state-of-the-art.

61. To search for information about a non-familiar subject

The interviewee considers that one of the possible tasks in the information-seeking process is to search for information about a subject that is not well known (or unknown).

When do you need to seek for information?
At the beginning, I need to find very basic papers that allow me to broadly understand

what the topic is about, without too much detail.

62. To cover a specific information need

The interviewee considers that one of the possible tasks in the information-seeking process is to answer to a specific information need that he/she may have. For example, he/she may need to search for an appropriate term to use in a specific domain, or he/she may need to obtain the definition of a term.

> *When do you need to seek for information?*
> *After writing the first draft of the paper, I send it to the other team members so they can identify deficiencies. If they do, it is necessary to perform very concrete searches in order to solve these deficiencies.*

63. To find again an article obtained in the past

The interviewee considers that one of the possible tasks in the information-seeking process is to search an article that he/she has previously found but that he/she is unable to remember exactly where or how to locate it.

> *When do you need to seek for information?*
> *When I want to find again a specific paper, I try to remember which were the keywords or ideas that led me to find it in order to use them again. Usually I also use the "page already visited" that indicates Google, but most of the times it is very difficult to find it again.*

64. To search for a target venue

The interviewee considers that one of the possible tasks in the information-seeking process is to search for an appropriate venue (conference, journal...) where his/her work can be submitted for publication.

> *When do you need to seek for information?*
> *First of all I need to find where I can publish my paper.*

65. To search for references for a cite

The interviewee considers that one of the possible tasks in the information-seeking process is to search for a reference where something specific is mentioned.

> *When do you need to seek for information?*
> *In other cases, there may be searches needed to add citations that support an idea or theory, mainly in the introduction. It's one of the most difficult tasks as you look for something very specific.*

66. To search for references for the state of the art

The interviewee considers that one of the possible tasks in the information-seeking process is to search for the articles that can serve as a basis for the description of the state of the art.

> *When do you need to seek for information?*
> *When I need to put my problem in a more general context: the problem is important, it hasn't been solved yet, it is based on proposals from other authors...*

67. To search for references that identify an unsolved need

The interviewee considers that one of the possible tasks in the information-seeking process is to search for references where a specific need is identified. For example, he/she may need to justify the need for his/her research work.

> *When do you need to seek for information?*
> *The first case is when I need to find related literature to be sure that what I want to publish hasn't been already published, that there is nothing similar and that it is novel.*

IDENTIFIED DIFFICULTIES
Search

68. To choose keywords

The interviewee mentions that he/she has found difficulties for choosing search keywords without explicitly specifying the reasons.

> *Which are the difficulties you have to face during the information-seeking process?*
> *It's difficult to become familiar with the correct terminology in order to elaborate queries that return the desired results, especially if I'm seeking for papers related with a topic I don't know.*

69. To choose non-ambiguous keywords

The interviewee mentions that he/she has found difficulties for choosing search keywords so that ambiguity is avoided. That is, results from other domains are obtained because of shared terminology.

> *Which are the difficulties you have to face during the information-seeking process?*
> *For example, when I do an undirected search and I feel a little bit lost, I use a lot of terms to construct the query, as I don't know if the terms I would use are correct or if there are other terms or expressions that would define better what I'm looking for. This implies that I have to refine the query based on what I read until I find the appropriate keywords.*

70. To choose keywords so that exhaustiveness is attained

The interviewee mentions that he/she has found difficulties for choosing search keywords in a

way that exhaustiveness is attained. That is, desired results have a very specific terminology, possibly unknown to the interviewee.

Which are the difficulties you have to face during the information-seeking process?
One of the difficulties is that it's not always easy to find the perfect query so that the search is exhaustive.

71. <u>To find again an article obtained in the past</u>
The interviewee mentions that he/she has found difficulties for finding again articles that he/she had previously consulted, analysed or read.

Which are the difficulties you have to face during the information-seeking process?
One of the biggest problems I have is that in many cases I don't remember where I have read a specific statement or which article covers a particular subject.

72. <u>Time required to obtain relevant articles</u>
The interviewee mentions as an obstacle the amount of time spent in trying to find and select articles relevant for his/her research.

Which are the difficulties you have to face during the information-seeking process?
When I look for papers dealing with a subject unknown for me, the process is more complex as it is longer and more tedious. It makes me spend a lot longer analysing papers in order to determine if they are of my interest.

73. <u>To know when the obtained results provide enough exhaustiveness</u>
The interviewee mentions that he/she has found difficulties to decide when the results obtained provide enough exhaustiveness, and thus the search process can stop.

Which are the difficulties you have to face during the information-seeking process?
The search relies too much on intuition and then I find it difficult to determine when I can consider the search as done because it has been systematic and exhaustive.

74. <u>To obtain complete access to certain articles</u>
The interviewee mentions as a difficulty to be able to obtain the full content of certain articles.

Which are the difficulties you have to face during the information-seeking process?
In scientific repositories, the main problem is that some articles are not accessible.

75. <u>To find an article including a specific assertion</u>
The interviewee mentions that he/she has found difficulties when trying to find an article containing a specific statement. For example, when he/she wants to find references to support the need of a specific research he/she is doing.

Which are the difficulties you have to face during the information-seeking process?
It is difficult to find somebody who says exactly what I want, moreover if the subject is not related with my research topic. In these cases, luck plays an important role.

Exploration

76. <u>High number of results</u>
The interviewee mentions as a difficulty for the exploration the high number of results obtained from the search.

Which are the difficulties you have to face during the information-seeking process?
Search engines return too many results.

77. <u>To know what has already been explored</u>
The interviewee mentions that he/she has found difficulties for knowing what he/she had previously explored, read or discarded.

Which are the difficulties you have to face during the information-seeking process?
One of biggest problems I usually have is that I lose the track of the things I have already read.

78. <u>High number of non-relevant results</u>
The interviewee mentions as a difficulty that a great number of results are not relevant for his/her research.

Which are the difficulties you have to face during the information-seeking process?
It is difficult for me to know which results are interesting, because there are many things that are worthless and it's difficult to find between them what I want.

79. <u>To establish if an article is a classical or an obsolete one</u>
The interviewee mentions as a difficulty to be able to decide if a given non-recent reference is either a classical or an obsolete one.

Which are the difficulties you have to face during the information-seeking process?
The age of the information can also be a problem because it is too old. In other words, it is difficult to identify which things are good and are not deciduous.

80. <u>To classify the articles</u>
The interviewee mentions as a difficulty to establish a good classification for the articles he/she selects as relevant.

Which are the difficulties you have to face during the information-seeking process?
Once I find a paper that I consider useful, I think it's difficult to classify it.

81. <u>The exploration is based on intuition</u>
The interviewee mentions as a difficulty the fact that the exploration of the results of a search depends on his/her intuition.

Which are the difficulties you have to face during the information-seeking process?
Exploring information is based on intuition, so if you don't have intuition it's very easy to stray from the subject and after this it's really difficult to come back.

Collaborative work

82 To know what has already been explored by other team members
The interviewee mentions that he/she has found difficulties for knowing what other members of his/her team had previously explored, read or discarded.
Are you part of a team where the information-seeking process is performed collaboratively? If so, does this affect the information-seeking process? How?
It is really difficult to make a collaborative search because you cannot know what articles have already consulted or at least considered the other members of the teamwork until they tell you. The problem is that this is very time-consuming and it's not worthy, so at the end we perform disjoint searches and then put them all together.

INFORMATION-SEEKING PROCESS

83. Iterative
The interviewee considers that the information-seeking process has an iterative nature with successive refinement.
How do you perform the information-seeking process?
I perform the process in several phases, refining more and more. This is due to the fact that I fear not considering things that can be important, and this makes me analyse a lot more than what I actually need. Therefore, I need to perform several levels of filtering in order to identify the papers that are really interesting for me.

84. Speed is important
The interviewee considers speed as an important characteristic for the information-seeking process.
How do you perform the information-seeking process?
I discard articles very quickly and I know that I probably let relevant things along the way, but I prefer to be fast and precise - as all the articles I keep are relevant – instead of exhaustive.

85. Precision is important
The interviewee considers precision as an important characteristic for the information-seeking process.
How do you perform the information-seeking process?
I discard articles very quickly and I know that I probably let relevant things along the way, but I prefer to be fast and precise - as all the articles I keep are relevant - instead of exhaustive.

86. Completion of every process stage ensures finding all relevant articles
The interviewee considers that following the whole information-seeking process ensures finding all relevant articles.
How do you perform the information-seeking process?
I skim a lot, paying attention only to things that attract my attention. By doing this, I probably let relevant things along the way, but I think that the process itself will take me to these important things later.

Variables affecting the process

87. User experience
The interviewee considers that the way the information-seeking process is carried out depends on how much experience the user possesses.
Has your information-seeking process evolved over the time? If so, why and how?
Of course it has changed, because while I have acquired experience, I have been refining and improving the procedure. It's an evolutionary process.

88. Available technology
The interviewee considers that the way the information-seeking process is carried out depends on the available technology.
Has your information-seeking process evolved over the time? If so, why and how?
It has changed a lot, especially because the technology has also evolved a lot.

89. User knowledge about the subject
The interviewee considers that the way the information-seeking process is carried out depends on how much the user knows about the subject at hand.
How do you perform the information-seeking process?
It depends on if I know the topic I'm investigating. If I do, I use specific sources of information to perform the search; if I don't, I search in Google or in the database of my University's library.

90. Working in a team
The interviewee considers that the way the information-seeking process is carried out varies when it is performed collaboratively.
Are you part of a team where the information-seeking process is performed

collaboratively? If so, does this affect the information-seeking process? How?
I work in a team with other researchers, and the main problem is the communication between us, which is not always easy and efficient. This can affect the whole process as, for example, it is difficult to know if any of the other members has already analysed or is analysing any of the papers returned as a result to my searches.

91. Spatio-temporal context

The interviewee considers that the way the information-seeking process is carried out depends on the spatio-temporal context in which the user is. This means that the moment when he/she seeks for information (e.g. early in the morning, late in the afternoon, Monday, Friday...) and the location where he/she does it (e.g. at the workplace, at home...) affects the information-seeking process.

Do you think the information-seeking process is always the same? Explain your answer.
The information-seeking process is never the same because each individual, both in space and in time, performs it differently. There are differences not only between different people, but also between the same person according to the moment, the context, the purpose... In brief, it depends on the user in the space-time.

92. Purpose

The interviewee considers that the way the information-seeking process is carried out depends on the specific aim he/she wants to achieve through the information-seeking process.

How do you perform the information-seeking process?
The search process depends on many aspects: the user, the available time, the purpose of the search, the expected type of results...

93. Available time

The interviewee considers that the way the information-seeking process is carried out varies according to the available time of the user.

How do you perform the information-seeking process?
I always do a first filtering using keywords in the query, and then I analyse the results more or less exhaustively according to the time available.

94. User

The interviewee considers that the way the information-seeking process is carried out depends on the user.

Do you think the information-seeking process is always the same? Explain your answer.
The search process depends on many aspects: the user, the available time, the purpose of the search, the expected type of results...

Variables not affecting the process

95, Purpose

The interviewee considers that the way the information-seeking process is carried out is the same regardless of the specific aim he/she wants to achieve through the information-seeking process.

Do you think the information-seeking process is always the same? Explain your answer.
I think the process is always the same, regardless of the purpose of the search. The only difference is, maybe, the amount of information I explore, that is how exhaustive is my exploration.

96. User knowledge about the subject

The interviewee considers that the way the information-seeking process is carried out is the same regardless of how much the user knows about the subject at hand.

Do you think the information-seeking process is always the same? Explain your answer.
I think that the experience of the researcher - more than the knowledge he/she can have about the topic - influences the way he/she searches and explores the information.

97. Working in a team

The interviewee considers that the way the information-seeking process is carried out is the same regardless of whether it is performed collaboratively or not.

Are you part of a team where the information-seeking process is performed collaboratively? If so, does this affect the information-seeking process? How?
In our group, looking for the literature we need is a team work, but the process doesn't change as we divide the work. That is, we perform individual searches and then we put them all together. Moreover, we communicate in person, face to face, so the process itself is the same.

Search subprocess

98. Focused search

The interviewee mentions that he/she performs a search with some initial orientation. This means that, for example, he/she knows what he/she wants to find, how to find it or even where to look for it.

How do you perform the information-seeking process?
At first, I directly consult the information sources - like journals o conferences - that I know that deal with my topic in order to see what has been made during the last months or years.

99. Free search
The interviewee mentions that he/she performs a free search. This means, for example, that he/she may not know exactly what he/she is looking for, where to look for it or how to find it.
> *How do you perform the information-seeking process?*
> *Initially, I perform an open search in the web, without any specific formulation, using terms that define, according to me, what I'm looking for.*

Exploration subprocess
100. Exhaustive exploration of the collection
The interviewee mentions that he/she makes a thorough exploration of every article. This means that he/she analyses every article in more detail.
> *How do you perform the information-seeking process?*
> *I know by heart most of the authors related with my topic, but nonetheless I don't trust myself and I perform exhaustive searches to verify that there is nobody else who has said something important or novel.*

101. Non-exhaustive exploration of the collection
The interviewee mentions that he/she does not explore the collection exhaustively. This means that he/she may not analyse every single article, or that the analysis made is very shallow.
> *How do you perform the information-seeking process?*
> *I skim a lot, paying attention only to things that attract my attention. By doing this, I probably let relevant things along the way, but I think that the process itself will take me to these important things later.*

102. Negative filtering
The interviewee mentions that he/she discards the articles that are not relevant for his/her research in the exploration process.
> *How do you perform the information-seeking process?*
> *Among the results of the searches that I perform in scientific repositories, I discard the articles that I consider useless according to their title.*

103. Positive filtering
The interviewee mentions that he/she selects the articles that relevant for his/her research in the exploration process.
> *How do you perform the information-seeking process?*
> *I analyse the results of a search sequentially. I take a look at the abstracts and store the ones I consider potentially interesting.*

META-INFORMATION
104. Wished after selecting an article
The interviewee mentions that he/she would like to obtain some meta-informaton but after the information-seeking process. This means that, once he/she she would have found an article and would have determined it is relevant for his/her research, he/she could use some meta-informaton for later purposes.
> *Do you miss something during the information-seeking process? If so, what?*
> *Once I select a paper because I think it can be useful, it would be interesting to know who has cited the paper or even the comments or ratings given by other researchers. However, I would only use this information after choosing the paper, as I want to decide only by myself if I keep or not a paper, without external influences.*

Used during search subprocess
105. Depends on the task at hand
The interviewee mentions that the meta-informaton he/she uses to perform a search depends on the tasks he/she needs to carry out.
> *How do you perform the information-seeking process?*
> *The search criteria vary according to the purpose. For example, if the reference papers I have cited in the article I'm writing seem to be obsolete, I look for new references filtering by year, whereas if there is a part of the article that is too superfluous or inaccurate, I seek per author or publication venue.*

106. References' authors of a relevant article
The interviewee mentions that he/she uses the name of the authors that appear in the references section of a relevant article to perform a search, for example including them in the query.
> *How do you perform the information-seeking process?*
> *If I think a paper is good and interesting, I use its references to continue the search: the journal where it has been published, the bibliography and the authors who have been cited.*

107. References' titles of a relevant article
The interviewee mentions that he/she uses the title of the references of a relevant article to perform a search, for example including them in the query.
> *How do you perform the information-seeking process?*
> *When I find a very interesting paper, I read the titles of its references, in case any of*

them could be also interesting.

108. Self-defined keywords

The interviewee mentions that he/she uses keywords he/she thinks are related with the topic he/she is researching in order to perform a search.

How do you perform the information-seeking process?
Initially, I perform an open search in the web, without any specific formulation, using terms that define, according to me, what I'm looking for.

109. meta-informaton extracted from a reference article

The interviewee mentions that the meta-informaton he/she uses to perform a search is obtained from an article he/she considers relevant for his/her research.

How do you perform the information-seeking process?
During the process, I use the bibliography and the authors of the most cited paper to refine my query.

110. meta-informaton known by heart

The interviewee mentions that he/she knows by heart the meta-informaton he/she uses to perform a search.

How do you perform the information-seeking process?
I search using the year, the authors or the keywords I can remember.

111. Authors

The interviewee mentions that he/she uses the name of one or more authors to perform a search, for example including them in the query.

How do you perform the information-seeking process?
The search criteria vary according to the purpose. For example, if the reference papers I have cited in the article I'm writing have become obsolete, I seek new references filtering by year, whereas if there is a part of the article that is too superfluous or inaccurate, I seek per author or publication venue.

112. Subject(s)

The interviewee mentions that he/she uses the subject(s) present in a relevant article to perform a search, for example including them in the query.

How do you perform the information-seeking process?
Sometimes I also perform searches in my local repository, restricting it to the folders related to the specific topic I'm working on.

113. Date of publication

The interviewee mentions that he/she uses a date as a filter to perform a search, for example excluding all the articles published before a specific year.

How do you perform the information-seeking process?
Now that I am an expert in my topic, the procedure has changed because I don't use generic repositories anymore, but I consult directly the journals that are relevant in my area, and I read all the latest issues.

114. Publication venue

The interviewee mentions that he/she uses the name of a specific venue as a filter to perform a search, for example excluding all the articles that do not belong to the already said venue.

Do you think the information-seeking process is always the same? Explain your answer.
How do you perform the information-seeking process?
Now that I am an expert in my topic, the procedure has changed because I don't use generic repositories anymore, but I consult directly the journals that are relevant in my area, and I read all the latest issues.

Used during exploration subprocess

115. Depends on the task at hand

The interviewee mentions that the meta-informaton he/she uses to decide if he/she keeps or rejects an article depends on the tasks he/she needs to carry out.

What information do you have into account when deciding whether to select an item?
It depends on the purpose of the search. For example, if I want to add reference papers in a Ph.D. thesis, I don't look at the type of the paper, because I want everything to be present. However, if I want to write a paper, I only take into account long papers, as I can't cite short papers.

116. Article's title

The interviewee mentions that he/she takes into account the title of an article to decide if he/she keeps it or rejects it.

What information do you have into account when deciding whether to select an item?
The title is very informative for me. I am able to obtain a lot of information only by reading it.

117. Article's authors

The interviewee mentions that he/she takes into account the authors of an article to decide if he/she keeps it or rejects it.

What information do you have into account when deciding whether to select an item?
Mainly I look at the title, the authors, the keywords, the abstract and the venue where it

has been published.

118. <u>Affiliation of the article's authors</u>

The interviewee mentions that he/she takes into account the affiliation of the authors of an article to decide if he/she keeps it or rejects it.

> *What information do you have into account when deciding whether to select an item? Moreover, if it has to do with a topic I know, I also take a look at the authors or their affiliation, as I know who is researching on the topic.*

119. <u>Article's publication date</u>

The interviewee mentions that he/she takes into account when was published an article to decide if he/she keeps it or rejects it.

> *What information do you have into account when deciding whether to select an item? In many cases, I also look at when and where has been published the paper, because if it is included in a good journal or conference, with a high impact, it is probable that the paper will also be good.*

120. <u>Impact of the article's publication venue</u>

The interviewee mentions that he/she takes into account the impact (e.g. Impact Factor in the JCR Report from Thomson Reuters) of the venue where an article has been published to decide if he/she keeps it or rejects it.

> *What information do you have into account when deciding whether to select an item? In many cases, I also look at when and where has been published the paper, because if it is included in a good journal or conference, with a high impact, it is probable that the paper will also be good.*

121. <u>Article's publication venue</u>

The interviewee mentions that he/she takes into account the venue where an article has been published (Conference, Journal...) to decide if he/she keeps it or rejects it.

> *What information do you have into account when deciding whether to select an item? Mainly I look at the title, the authors, the keywords, the abstract and the venue where it has been published.*

122. <u>Article's keywords</u>

The interviewee mentions that he/she takes into account, if available, the keywords of an article to decide if he/she keeps it or rejects it.

> *What information do you have into account when deciding whether to select an item? Mainly I look at the title, the authors, the keywords, the abstract and the venue where it has been published.*

123. <u>Article's language (only English)</u>

The interviewee mentions that if an article is not written in English, he/she rejects it directly.

> *What information do you have into account when deciding whether to select an item? If I want to use it as a reference in a paper I'm writing, I only consider papers written in English.*

124. <u>Article's language (any other known language)</u>

The interviewee mentions that he/she does not take into account in which language is written an article to decide if he/she keeps it or rejects it, as long as he/she can understand it.

> *What information do you have into account when deciding whether to select an item? Usually I don't take into consideration the language of the article, as long as it is a language I can understand. In fact, on occasion, I have found papers that seemed very interesting in a language that I didn't understand, and I managed to translate them in order to read them.*

125. <u>Article's length</u>

The interviewee mentions that he/she takes into account how many pages has an article to decide if he/she keeps it or rejects it.

> *What information do you have into account when deciding whether to select an item? Short papers usually present a preliminary work, and then they are not useful for some tasks, like preparing the state-of-the-art of a paper, so I directly discard them.*

126. <u>Article's impact</u>

The interviewee mentions that he/she takes into account, if available, how many people have cited an article to decide if he/she keeps it or rejects it.

> *What information do you have into account when deciding whether to select an item? It is important for me that many other articles have cited an article as this indicates that the latter has been well written and provides a significant contribution.*

127. <u>Article's type</u>

The interviewee mentions that he/she takes into account what type is an article (Review, Study, Divulgation...) to decide if he/she keeps it or rejects it.

> *What information do you have into account when deciding whether to select an item? The sifting is made based on the quality of the information and the purpose of the paper: if I want my students to read it, the paper has to be mainly informative, whereas if I want to it to write a paper, it has to be deeper and provide significant results.*

128. <u>Article's structure</u>

The interviewee mentions that he/she takes into account the structure of an article (present sections and their order) to decide if he/she keeps it or rejects it.

> *What information do you have into account when deciding whether to select an item?*
> *I like to verify that the structure of the paper is logical and if it fits what I'm looking for.*

129. Article's abstract

The interviewee mentions that he/she reads to some extent the abstract of an article to decide if he/she keeps it or rejects it.

> *What information do you have into account when deciding whether to select an item?*
> *Mainly I look at the title, the authors, the keywords, the abstract and the venue where it has been published.*

130. Article's introduction section

The interviewee mentions that he/she reads to some extent the introduction of an article to decide if he/she keeps it or rejects it.

> *What information do you have into account when deciding whether to select an item?*
> *It depends on the objective. If want it for the state-of-the-art section I have to perfectly understand the proposed solution, whereas if I want to know more about a topic, I focus more on the introduction section.*

131. Article's conclusions section

The interviewee mentions that he/she reads to some extent the section of conclusions of an article to decide if he/she keeps it or rejects it.

> *What information do you have into account when deciding whether to select an item?*
> *First of all I read the title, and if I like it, then I take a look at the abstract. If it matches more or less what I am looking for, then I read the conclusion.*

132. Authors of the article's references

The interviewee mentions that he/she takes into account the authors of the articles referenced in the bibliography of an article to decide if he/she keeps it or rejects it.

> *What information do you have into account when deciding whether to select an item?*
> *One of the things I consider is if the authors that are leaders in the topic have been cited, because if so, it is more probable that the paper has a good basis and fits with my topic.*

133. Amount of references in the article

The interviewee mentions that he/she takes into account how many references has an article in its bibliography section to decide if he/she keeps it or rejects it.

> *What information do you have into account when deciding whether to select an item?*
> *I consider which sections has the paper, how many pages it has, where it has been published, how many references are in the bibliography, which authors has been used in the bibliography, age of the references...*

134. Publication date of article's references

The interviewee mentions that he/she takes into account when were published the references included in the bibliography section of an article to decide if he/she keeps it or rejects it.

> *What information do you have into account when deciding whether to select an item?*
> *I consider which sections has the paper, how many pages it has, where it has been published, how many references are in the bibliography, which authors has been used in the bibliography, age of the references...*

135. Existence of graphical elements

The interviewee mentions that he/she takes into account if an article has graphical elements (e.g. Illustrations, Graphs, Schemes...) to decide if he/she keeps it or rejects it.

> *What information do you have into account when deciding whether to select an item?*
> *Sometimes I take a look at the pictures or graphs that the paper may contain. In fact, if these illustrations are excellent or very novel, I use them as examples to illustrate my articles, especially when I don't know how to represent some things.*

Wished during exploration subprocess

136. Article's most frequent terms

The interviewee mentions that he/she would like to obtain the list of terms of an article ranked according to their frequency of appearance to decide if he/she keeps it or rejects it.

> *Do you miss something during the information-seeking process? If so, what?*
> *It would be interesting to know which are the most frequent terms in a paper, so that I could see if what I am looking for has been mentioned many times.*

137. Ranking of article's keywords

The interviewee mentions that he/she would like to obtain the list of keywords of an article ranked according to their relevance in the article to decide if he/she keeps it or rejects it.

> *Do you miss something during the information-seeking process? If so, what?*
> *I would like the keywords of the articles were ranked according to their importance and to their weight in the paper.*

138. Article's concepts

The interviewee mentions that he/she would like to know what concepts are addressed in an article, even if they do not appear explicitly written, to decide if he/she keeps it or rejects it.

> *Do you miss something during the information-seeking process? If so, what?*

Mainly I would love to know what are the general concepts addressed in a paper, because this would spare me from having to read the title, the abstract, the conclusions... The concepts would suffice to determine if a paper is interesting or not.

139. <u>Authors who have cited the article</u>
The interviewee mentions that he/she would like to know which authors have cited an article to decide if he/she keeps it or rejects it.
Do you miss something during the information-seeking process? If so, what?
For other research tasks different from writing papers, like finding reviewers for a paper or committee members for a conference, it could be useful to know what authors have cited a paper, the impact of the authors of a paper, or the impact of the authors of the bibliographic references of a paper in order to know who is working in the topic and who is an expert in the topic.

140. <u>Authors of the article's references</u>
The interviewee mentions that he/she would like to know which authors have been cited in the bibliography section of an article to decide if he/she keeps it or rejects it.
Do you miss something during the information-seeking process? If so, what?
I would like to see both the authors who have cited the article and those who have been cited in the article.

141. <u>Comments from other researchers</u>
The interviewee mentions that he/she would like to know the comments and scores that other researcher may have given to an article to decide if he/she keeps it or rejects it.
Do you miss something during the information-seeking process? If so, what?
I would like to have some kind of section with comments and ratings from other researchers who have already read the paper, like it is done in online retailers or in systems for online hotel reservations.

142. <u>Impact of the article's authors</u>
The interviewee mentions that he/she would like to know how many people have cited the authors of an article to decide if he/she keeps it or rejects it.
Do you miss something during the information-seeking process? If so, what?
For other research tasks different from writing papers, like finding reviewers for a paper or committee members for a conference, it could be useful to know what authors have cited a paper, the impact of the authors of a paper, or the impact of the authors of the bibliographic references of a paper in order to know who is working in the topic and who is an expert in the topic.

143. <u>Impact of the article's publication venue</u>
The interviewee mentions that he/she would like to know the impact (e.g. Impact Factor in the JCR Report from Thomson Reuters) of the venue where an article has been published to decide if he/she keeps it or rejects it.
Do you miss something during the information-seeking process? If so, what?
The system could indicate the impact factor of the journal where the paper has been published in the year when it has been published without needing to look for it manually.

144. <u>Impact of the article's references</u>
The interviewee mentions that he/she would like to know how many people have cited the references cited in an article to decide if he/she keeps it or rejects it.
Do you miss something during the information-seeking process? If so, what?
Also, it would be interesting to know how many times have been cited the papers used as references in a paper. This way I could know how much impact has its bibliography.

145. <u>Impact of the authors of the article's references</u>
The interviewee mentions that he/she would like to know how many people have cited the authors of the references cited in an article to decide if he/she keeps it or rejects it.
Do you miss something during the information-seeking process? If so, what?
For other research tasks different from writing papers, like finding reviewers for a paper or committee members for a conference, it could be useful to know what authors have cited a paper, the impact of the authors of a paper, or the impact of the authors of the bibliographic references of a paper in order to know who is working in the topic and who is an expert in the topic.

REFERENCE MANAGER
146. <u>Used</u>
The interviewee mentions that he/she uses a reference manager.
How do you store he articles once you have selected them?
I keep the references sorted by publication year in a bibtex file in CiteULike.

147. <u>Not used</u>
The interviewee mentions that he/she does not use a reference manager.
How do you store he articles once you have selected them?
I have never used a reference manager, because I don't write too many papers and then the bibliography I manage is not very extensive.

Usage

148. <u>To save articles</u>

The interviewee mentions that he/she uses the reference manager as a repository.

> *How is the reading process of the articles you consider relevant?*
> *I upload the file with my handwritten comments to CiteULike and add notes in the reference manager itself about the most relevant aspects of the paper.*

149. <u>To save annotations about articles</u>

The interviewee mentions that he/she adds annotations (comments, tags...) to articles in the reference manager.

> *How is the reading process of the articles you consider relevant?*
> *I upload the file with my handwritten comments to CiteULike and add notes in the reference manager itself about the most relevant aspects of the paper.*

150. <u>To perform searches on the stored bibliography</u>

The interviewee mentions that he/she uses the reference manager as a search engine for the documents managed in it.

> *How do you perform the information-seeking process?*
> *I also perform searches among the papers I have stored locally in my computer using the reference manager.*

READING PROCESS
Activities performed

151. <u>Underlining</u>

The interviewee mentions that he/she underlines the parts of the article he/she considers relevant (either on paper or in digital format).

> *How is the reading process of the articles you consider relevant?*
> *I read them more in detail, and I write down comments and highlight the most relevant parts.*

152. <u>Writing annotations in a separate document</u>

The interviewee mentions that he/she writes annotations (notes, summaries, key points of the article, ideas, authors...) during the reading, but on a separate document different from the article.

> *How is the reading process of the articles you consider relevant?*
> *While I read the article, I take notes in a separate file in Excel or Atlas.TI. I also comment the paper, both in the digital copy and in the printed one.*

153. <u>Writing annotations in articles</u>

The interviewee mentions that he/she writes annotations (notes, summaries, key points of the article, ideas, authors...) during the reading directly on the article itself.

> *How is the reading process of the articles you consider relevant?*
> *I read very actively: I write down notes, ideas, opinions... on the printed article itself while I read it.*

Outcome

154. <u>Elaboration of a mental map of reference authors</u>

The interviewee mentions that he/she creates and updates his/her mental map of reference authors while he/she is reading articles. The map can be only mental, but it also can be transcribed to a physical or digital document.

> *How is the reading process of the articles you consider relevant?*
> *I create a little local collection with the papers returned by the search and read them more in detail, paying particular attention to the state of the art and the bibliography, identifying who are the authors that are referenced. This allows me to identify and group the most cited authors and papers.*

155. <u>Elaboration of a mental map of reference articles</u>

The interviewee mentions that he/she creates and updates his/her mental map of reference articles while he/she is reading articles. The map can be only mental, but it also can be transcribed to a physical or digital document.

> *How is the reading process of the articles you consider relevant?*
> *I create a little local collection with the papers returned by the search and read them more in detail, paying particular attention to the state of the art and the bibliography, identifying who are the authors that are referenced. This allows me to identify and group the most cited authors and papers.*

USER PROFILE
Expertise and knowledge

156<u>Use of expertise</u>

The interviewee mentions that the strategy he/she uses to seek the information steams from the experience he/she has obtained while doing previous information-seeking processes. For example, he/she can have a systematic methodology to select the keywords to construct the queries.

Has your information-seeking process evolved over the time? If so, why and how?
Over time I have become an expert in the topic I investigate, which allows me to assign more importance to my knowledge and to my experience than to the system, which I only use as a tool to fill specific gaps I can have.

157. No use of expertise
The interviewee mentions that he/she does not use his/her acquired expertise during the information-seeking process.
Has your information-seeking process evolved over the time? If so, why and how?
Personally, not really, because I don't consider myself an expert in anything, and then I have to start every time almost from scratch.

158. Use of own knowledge
The interviewee mentions that he/she uses his/her prior knowledge as a starting point or during the process of information-seeking. For example, he/she can directly select an article because he/she knows that one of the authors is relevant in the field of his/her research.
How do you perform the information-seeking process?
I use a lot my memory because I trust it a lot.

159. No use of own knowledge
The interviewee mentions that he/she does not use his/her prior knowledge during the process of information-seeking.
Has your information-seeking process evolved over the time? If so, why and how?
Not too much. I think I still don't have too much expertise, plus I don't have enough confidence in myself to trust my memory and my knowledge, and then I usually don't use them.

160. User has a mental map of reference authors
The interviewee mentions that he/she has a mental map of authors that are relevant in the subject he/she is researching. The map can be only mental, but it also can be transcribed to a physical or digital document.
How do you perform the information-seeking process?
I have some mental maps in my mind that I tend to use, for example authors or journals that are a relevant in my field of work.

161. User has a mental map of reference venues
The interviewee mentions that he/she has a mental map of venues (journals, conferences, workshops...) that are relevant in the subject he/she is researching. The map can be only mental, but it also can be transcribed to a physical or digital document.
How do you perform the information-seeking process?
At first, I directly consult the information sources - like journals o conferences - that I know that deal with my topic in order to see what has been made during the last months or years.

162. User has a mental map of reference articles
The interviewee mentions that he/she has a mental map of articles that are relevant in the subject he/she is researching. The map can be only mental, but it also can be transcribed to a physical or digital document.
How do you perform the information-seeking process?
I trust a lot the search engines and repositories because my memory fails. I have in mind many papers and authors, but I am aware that my memory has a limit, and then it is important for me to be able to perform a search to avoid forgetting any of them.

Personal preferences
163. Maintain full control
The interviewee mentions that he/she wants to be the only one who decides how the information-seeking process is performed. This means that he/she does not want the system to perform any task by itself (e.g. to apply filtering in the search process, to offer suggestions, to classify articles...).
How do you perform the information-seeking process?
The selection of a paper has to depend only on me, not on other external people or things

Personal characteristics
164. High self-confidence
The interviewee mentions that he/she trusts his/her judgment and feels confident making decisions during the information-seeking process.
Has your information-seeking process evolved over the time? If so, why and how?
Now I have a lot more confidence in myself as I have more knowledge and experience, and then I am able to select and discard papers potentially interesting with less information, more quickly and better than before.

165. Lack of self-confidence
The interviewee mentions that he/she does not trust too much his/her judgment or he/she feels unconfident making decisions during the information-seeking process.
Has your information-seeking process evolved over the time? If so, why and how?

Even today, I don't trust myself too much because I have read so many papers that I don't know where I have read what I need.

COLLABORATIVE WORK

166. <u>Information-seeking process is performed collaboratively</u>

The interviewee mentions that he/she works in a team, and the information-seeking process tasks are shared between team members.

How do you perform the information-seeking process?

We split the search results between the members of a workgroup. Then, as we read and select the papers, we write down a short summary of each of them in a shared document.

167. <u>Information-seeking process is not performed collaboratively</u>

The interviewee mentions that he/she works in a team, but the information- seeking process tasks are not shared between team members. Therefore, each member of the team performs the full information-seeking process individually.

Are you part of a team where the information-seeking process is performed collaboratively? If so, does this affect the information-seeking process? How?

Even if I work in a team, we perform a hierarchical division of non-concurrent tasks, thus we actually don't work collaboratively.

168. <u>The team uses a shared file for annotations</u>

The interviewee mentions that he/she works in a team, and that team members write annotations in a shared file, so the rest of the team can access them.

How is the reading process of the articles you consider relevant?

Each member of the team writes in a page shared with the rest of the team a summary of the papers he/she has read, its pros and cons, the most relevant sections to read, the main contributions...

169. <u>The team exchanges relevant articles</u>

The interviewee mentions that he/she works in a team, and that the team shares a pool of articles that may be relevant to some members.

How is the reading process of the articles you consider relevant?

When we think that a paper can be useful for another member of the team, we send it to him/her so he/she can take a look at it.

Contents | **Author index** | **Subject index** | **Search** | **Home**