The role of current working context in professional search

Maya Sappelli
TNO & Radboud University Nijmegen
m.sappelli@cs.ru.nl

ABSTRACT

The focus of the current research project is increasing or maintaining well-being at work through information support. The proposed research has two goals: (1) to optimize the user profile that is used to understand what the user wants and needs in terms of information support and (2) to evaluate the user model in the context of tools that assist knowledge workers in managing their information flow. One of the applications we study is professional search. We are currently preparing a series of experiments on improving professional search by incorporating the current context of the searcher. The context we would like to take into account is: the task the user is performing, the produced content so far, queries issued during the task and the information objects that were accessed during the task. The central questions are (1) how to represent a user model that incorporates the current working context and (2) how to use this model to improve professional search.

Categories and Subject Descriptors

H.3.4 [**Information Storage and Retrieval**]: Systems and Software—*Performance evaluation, User profiles and alert services*; I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Text analysis*

1. INTRODUCTION

This research project is part of a bigger project named SWELL[1] There are several industrial and research partners involved in the project. The overall objective is to monitor an individual in home and work settings and provide them with an unobtrusive coach or assistant to increase the individual's well-being.

Within SWELL, it is my proposal to focus on increasing or maintaining well-being at work through information support. For knowledge workers (i.e. people that use and produce knowledge) who work behind computers most of their time, keeping track of their activities and managing their information flow can become difficult. This is a result of the large amount of information that is processed and produced during a day. When information flow is not properly managed, people can feel unproductive or inefficient. The proposed research will have two goals. The first goal is to optimize the user profile, which should capture what the user wants and needs in terms of information support. The model can be used to determine which information is relevant for the user and his current activities, which can help filter the information that the user encounters without losing relevant or important information. The second goal is to evaluate the user model in the context of tools that assist knowledge workers in managing their information flow. We aim to do this in a small-scale field study. We focus on e-mail management and professional search as application areas for the user model.

In this doctoral consortium proposal we motivate our project and describe relevant literature including previous work. We then present our plans for a series of experiments that aim at improving professional search by incorporating the current context of the searcher. We hope to receive feedback on how to incorporate current context in a user model and how to improve search results by taking the current context into account. As current context we will use the following types of information: the task the user is performing (e.g. writing a report, or preparing a presentation), the produced content so far, queries issued during the task and the information objects that were accessed during the task. In addition, we would like to receive feedback on how we could take temporal aspects of user activity into account (e.g. when is someone executing a task, how do the user's interests change over time).

2. BACKGROUND AND RELATED WORK

There are several areas of research that are associated with the presented work. First, research on information overload, well-being and productivity in a work setting is relevant for the motivation of this research and the overall SWELL project. Second, research on user profiling methods and term extraction are important to understand what the limitations are of current user profiles. In SWELL, context has been identified as the main solution area to determine what a user needs. Therefore, research into recognizing context in the work domain is relevant as well. Finally, we look at research on email categorization and personalized professional search as we will focus on these research areas as our main evaluation platforms.

[1]http://www.swell-project.net

SIGIR 2013, Dublin, Ireland

2.1 Well-being at work

In 1990, Schick et al. [27] defined information overload as "occurring when the information processing demands on an individual's time to perform interactions and internal calculations exceed the supply or capacity of time available for such processing". It is believed that this reduces an individual's decision making capabilities. The authors state that information overload can be defined by the quantity of information that needs to be processed per unit of time. They believed that information overload can be reduced by making more efficient use of time (for example by standardizing operations, training or reducing the number of tasks that need to be performed) or increasing the time available in organizations (for example by increasing the time available for each individual or by expanding the workforce).

Ruff [22] performed a review study on the effects of information overload on performance, physical health and social relations and found that more than 60% of the employees reported a negative impact of information overload on personal and collegial relationships. Additionally, more than 25% of workers experience stress or health issues caused by information overload. Effects of information overload include concentration problems, multitasking, or compulsion to check email and internet. Solutions for information overload are according to Ruff et al.: (1) filtering (focusing attention on the most important information), (2) escaping (limiting disruptions from the outside world), (3) prioritizing tasks, satisficing (use "good-enough" rather than "perfect" solutions) and (4) limiting (accepting that more information is not always better). Spira and Goldes [30] add the necessity to think before you send an e-mail to the list of information mitigation strategies.

Many automatic solutions to filtering information such as categorizing email messages or documents have been suggested. However, when viewed from a psychological perspective [19], Lansdale notes that there is a general problem with categorising items, caused by the cognitive process of recall. For people, it is difficult to determine a categorisation for items, but it is also difficult to remember what labels they used for te categorisation, thereby making it difficult to re-find items. This is because most information items do not fall into neat categorisation structures and category names can be ambiguous. Additionally, we remember much more about documents and the context of a document (for example where you were sitting while you wrote the document, or what it looked liked) then an automated retrieval process uses. Therefore, according to Lansdale, information retrieval and filtering techniques should be based on cognitive recall, meaning that they should take into account how the user remembers information, and not only how the computer can find it.

2.2 User modelling

According to the literature described in the previous section, knowledge workers would be helped by solutions targeted at reducing the time necessary to find information and methods that support the organization of information. For that purpose, we need to understand what information a knowledge worker needs. Our previous research into predicting the intent behind a user's query to a search engine [26] revealed that even if the user formulates an explicit query, this alone is not enough to assess what the user is looking for. The context that lead to the query can be a valuable source of information to better understand the user's needs, as well as information about the user's interests, expertise, query history etc. Additionally, Gomez-Perez et al. [11] suggest that knowledge workers can benefit from working in context, a method of working in which information objects are associated with a context (e.g. objects that are frequently accessed concurrently or are similar in content) such that incoming information sources that belong to other contexts can be filtered out to make sure that the knowledge worker remains focused on the relevant and important sources.

In general there are 3 types of user models [13, 9]: 1) static, in which main data is gathered and is not updated 2) dynamic, in which changes in interests or interactions are recorded and influence the model 3) stereotype based, in which a group model is used. The main approaches to acquire the user models are by a) asking the user, 2) observing and interpreting user interaction or 3) hybrid method in which users are asked first and the model is adapted after observing the user, by for example relevance feedback.

In Information Retrieval, the user can be modelled in terms of his information need. The user is asked to formulate a query, specify a query or reformulate a query. This can be seen as a form of a static adaptable user model elicited from the user. More extensive user models in information retrieval also model the user's background by incorporating the query history or some elicited goals or domain knowledge, making them more adaptive. Relevance feedback is a method to observe and interpret the user and can be used to adapt the model. Overall the user model is often limited to include information on previous interactions with the search system. [10]

Shen et al. [29] present a decision theory framework for implicit user modelling for IR. They model the user in terms of information need, previously viewed documents and interaction history.

Abel et al. [1] compare hashtag-based, entity-based and topic-based user profiles for Twitter. They enrich them with entities and topics from linked news articles and also investigate temporal effects in the profiles. Differences between the profiles for the week and for the weekend days were found, but tweets alone were not sufficient to understand the variation in user's concerns and interests. Entity-based and topic-based user profiles showed advantages over the hashtag based profiles.The addition of linked news articles enhanced the variety of the profiles and improved the accuracy when recommending news articles.

In our project we aim to gather more extensive knowledge about the user. User interests, expertise, social relations and current work-related activities (e.g. the project or topic a user is working on at this moment) are seen as important parts of the model. These elements need to be recognized automatically from the information objects accessed by the user, and not only the interactions with the search system. Overall, this means that we need a highly adaptive user model acquired using a hybrid method of observing the user and asking the user for feedback on the model directly.

2.3 Context recognition

Context is a key factor to help filter information objects and should be an important part of the user model as well. Another way to look at context is to see it as a partial identity. Each project or topic that a user works on has its own dimensions For example, each project has its own members associated with it, documents that are created or read etc. Of course partial identities can overlap. Recognizing a context can be seen as recognizing which partial identity is active at the current moment. This partial identity has a relation to information objects and other individuals. However, a context can be broader than a partial identity and can include for example which type of activity the user is doing (e.g. reading, writing, preparing presentation) or the current location.

There are several approaches to recognizing contexts. Warren et al. [34] distinguish in their ACTIVE project the process of context discovery and the process of context detection. In the first process new contexts are recognized (based on content similarities, but also based on co-occurrence of document access), while in the latter the current context is identified. Both processes are automated using machine learning techniques. In context discovery, streams of events are captured. Contents of documents, webpages and e-mails accessed in these streams of events are analysed and clusters of information sources, collaborators and tasks are identified. The users have to interpret the suggested contexts, give them names and adapt them when necessary. In the context detection phase, streams of events are mapped to the identified contexts. There are explicit associations between information sources and contexts, since the information sources are tagged with the context that was active when the user accessed the document. When the system recognizes that an information source belongs to another context, because that source is tagged with another context, the system notifies the user of a possible context switch. The user can then initiate the context switches when he feels it is appropriate, or tag the information source with the current context. The approach taken in the ACTIVE project still requires much user effort since contexts are not defined or switched without user input.

Other approaches to context recognition include the work by Shen et al. [28], Granitzer et al. [12], Kellar and Watters [15], Bauer et al. [4], Cheyer et al. [5] and Oliver et al. [21]. The interpretation of context varies but are related to knowledge workers task activities (e.g. writing a report) and task content (topic X, project Y).

Koldijk et al. [18] collected labelled examples of task activities using a key logger to monitor a knowledge worker's activity. They asked users on regular intervals to label the task activity they were doing. Intuitive task labels were acquired in cooperation with some users and included: reading or writing e-mail, writing a report, programming, analysing data and searching for information. Additionally, they investigated whether these tasks can be recognized automatically from the low level log events (such as mouse or key activity) using automated classifiers. They found that with relatively little labelled data, i.e. a few hours worth, reasonable classification accuracy could be obtained. However, there were many individual differences and there was no single classifier that performed consistently over all users.

In our approach, context will consist of the recognized task [18], the recognized topic/project, and the social relations (e.g. collaborators, clients). We propose to recognize these elements using an activity logger that keeps track of user interactions with the computer (e.g. document or URL accessed, queries, application started/stopped). This approach is similar to the ACTIVE approach [34]. ACTIVE, however, does not monitor low-level events such as keystrokes, mouse clicks or window selections. Additionally, ACTIVE requires the user to initiate context switches and confirm recognized contexts, which we would preferably avoid. Furthermore, we will look into the recognition of context using Screen OCR, which gives us more textual and content data, but also more noise.

2.4 Application areas for the knowledge worker's user model

2.4.1 E-mail management

A number of applications for e-mail management, such as spam-detection [23] , prioritization [2], folder prediction [16, 20] and mail categorization [14, 7, 16] have been described in the literature.

In previous work [24], we have investigated whether folder structure and documents could be exploited as labelled examples to categorize mail. The purpose of this experiment was to reduce the input necessary from the user and to improve the coupling between documents and email messages.

However, in all these methods the user is only taken into account by the trained examples he provides. We would prefer an unsupervised method, which minimizes the user's workload, but unsupervised methods often lack the user's point of view. In our project we want to be able to show the user which documents or e-mails are important and why these are relevant and important for him. Traditional categorization methods would describe a document in the same terms for each user. But imagine a historian that finds a document about democracy in 1800. The historian is working on some research about the era of 1790-1810. He might categorize this as "democracy". Imagine another user, a politician, that is working on the history of democracy. He would likely categorize the same document as "1800" rather than "democracy". This means that for user's to adopt and understand a proposed categorization, we need an unsupervised method that categorizes a document differently for different people.

Another application area is reply prediction (the task of predicting whether an e-mail message will be replied to), which gives an indication of how important a message is. In this type of research, the main personal information that is used, is the frequency of communication between a sender and a receiver of a message [8, 2, 3]. The topic of the message or whether it fits the current activities is not taken into account, even though this may influence whether the message is considered to be important or not.

2.4.2 Professional search

In the Information Retrieval field there are many examples of search approaches in which the user's history is taken into account (See [10] for an overview of Personalized IR).

Related to our objectives, Sugiyama et al. [31] adapted their search system to the users by modelling short term and long term interests and incorporating collaborative filtering techniques.The user profile constructed based on modified collaborative filtering achieved the best accuracy. This method used browser history and persistent terms, and added terms from the history of users that are similar to the active users.

Chirita et al. [6] proposed to expand short query keywords with terms from a user's personal information repository. They used 5 techniques for search result personalization: Term and Document Frequency, Lexical Compounds, Sentence Selection, Term Co-occurrence Statistics and Thesaurus Based Expansion. Term Frequency and Lexical Compounds performed best.

Verberne et al [33] attempted to incorporate background knowledge to improve search results in academic professional search. However, incorporation of a user model based on background knowledge did not improve results of a model based on the query history (merged model of the search terms).

However, there are still opportunities for improvement in these methods when it comes to using more diverse data from the user and combining it with long and short term interests.

3. DESCRIPTION OF RESEARCH PROJECT

The current project focuses on discovering what elements we need to properly understand what a user wants and needs in terms of information support and how we can use that profile to filter information. For this purpose we will develop new methods to describe a user. We will mainly use textual data to describe the user, but aim to mix this with other relevant data such as social relations. One challenge will be to use data from different (textual) sources with various characteristics, together with non-textual data to filter the information. The aim is to develop a user model that describes the user well, is intuitively recognizable by the user, and can be used to enhance existing IR or classification techniques. The user model will consist of several partial identities, that are related to the contexts the user works in. The questions that we aim to answer in the current project are:

- How do we create and organize a user model for knowledge workers that can be used for information filtering?
- How do we use the user model for information support?

For the evaluation of our user model we propose two approaches. In the first approach we assess the user model intrinsically. Our goal is to develop a transparent user model in which the user recognizes himself and which the user can adapt. For this purpose, we will do experiments in which we will ask colleagues to assess the relevance of the proposed user model [32]. Additionally, we ask them if they can tell which profile belongs to which colleague. The idea behind this is to distinguish the internal view from the external view. The internal view is important for the personalized filtering assistant, however the external view can be used for expert finding. Understanding the differences between the two may help to define the user model.

Secondly we aim to assess the model extrinsically by comparing various types of user models in their performance on the same task. We will mainly look at their performance on search tasks and email classification tasks (i.e. categorization, importance of the message, relevance of the message to the context). It is interesting to look at both types of tasks, since they both have different characteristics and may need different parts of the user model. We will focus on using various term extraction techniques and various combinations of features.

4. PROPOSED RESEARCH

Context is an important aspect in our project. Currently, we have a general idea on how to model and how to recognize this context. However, our question is how we can use this elaborate context to improve the retrieval process.

Recently, we have collected a dataset of user activities (i.e. the current context of the user) during typical knowledge worker's tasks [17]. We plan to use these data for a series of experiments in which we aim at improving professional search by taking the searcher's current context into account. The dataset consists of a full log of user activity on a computer. The participants were asked to write reports on a fixed set of 6 topics (tasks): a) opinion and facts on stress at work, b) opinion and facts healthy living, c) opinion and facts privacy on the internet, d) plan a coast-to-coast roadtrip in the USA, e) plan 5 tourist activities in Perth Australia, e) a short biography of the life of Napoleon

The participants were also asked to prepare presentations for 3 of the 6 topics and additionally they received email messages with another short task (either a request to send a picture on information overload, or a request for the birthdate of Einstein). In total, 25 students and interns participated. All participants searched on-line for the information they needed for completing the tasks. Table 1 presents an overview on the number of information objects and queries that are available in the dataset.

Table 1: Details on the available data

No. task-related queries	708
Average query length	3.39 words
Average no. queries per task	89
Average no. queries per participant	24
No. task-related information objects	4721
Average no. information objects per task	590
Average no. information objects per participant	330

From the recorded data we can observe which content context was active (e.g. on which topic their current report was), which queries they used, which websites were accessed, how long they stayed on that website, and whether they typed in their report after looking at a website. We plan to use the last two elements as relevance feedback. The central questions of our experiments are (1) how to represent a user model that incorporates the current working context and (2) how to use this model to improve professional search.

5. ISSUES FOR DISCUSSION

There are a number of open issues related to the planned experiments that we would like to discuss at the Doctoral Consortium meeting:

- How can we use the dwell time on information objects and switches back to the document that is being written as estimators for relevance?
- How could we take temporal aspects of user activity into account (e.g. when is someone executing a task, how do the user's interests change over time).
- How can we use the current context to improve search results? Previous work [33] on incorporating more information about a user showed that a simple merged query model (query history and background knowledge) is not sufficient to improve search results. Therefore we would like to discuss this experiment and how we can best incorporate this elaborate context.

6. ACKNOWLEDGEMENTS

This publication was supported by the Dutch national program COMMIT (project P7 SWELL).

7. REFERENCES

[1] F. Abel, Q. Gao, G.-J. Houben, and K. Tao. Analyzing user modeling on twitter for personalized news recommendations. *User Modeling, Adaption and Personalization*, pages 1–12, 2011.

[2] D. Aberdeen, O. Pacovsky, and A. Slater. The learning behind gmail priority inbox. In *LCCC: NIPS 2010 Workshop on Learning on Cores, Clusters and Clouds*, 2010.

[3] T. Ayodele and S. Zhou. Applying machine learning techniques for e-mail management: solution with intelligent e-mail reply prediction. *Journal of Engineering and Technology Research*, 1(7):143–151, 2009.

[4] T. Bauer and D. Leake. Word sieve: A method for Real-Time context extraction. *Modeling and Using Context*, pages 30–44, 2001.

[5] A. Cheyer, J. Park, and R. Giuli. Iris: Integrate. relate. infer. share. In *Semantic Desktop Workshop*. Citeseer, 2005.

[6] P.-A. Chirita, C. S. Firan, and W. Nejdl. Personalized query expansion for the web. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 7–14. ACM, 2007.

[7] E. Crawford, J. Kay, and E. McCreath. *Automatic Induction of Rules for e-mail Classification.*

[8] M. Dredze, T. Brooks, J. Carroll, J. Magarick, J. Blitzer, and F. Pereira. Intelligent email: Reply and attachment prediction. In *Proceedings of the 13th international conference on Intelligent user interfaces*, pages 321–324. ACM, 2008.

[9] G. Fischer. User modeling in human–computer interaction. *User modeling and user-adapted interaction*, 11(1):65–86, 2001.

[10] M. R. Ghorab, D. Zhou, A. O'Connor, and V. Wade. Personalised information retrieval: survey and classification. *User Modeling and User-Adapted Interaction*, pages 1–63, 2012.

[11] J. Gomez-Perez, M. Grobelnik, C. Ruiz, M. Tilly, and P. Warren. Using task context to achieve effective information delivery. In *Proceedings of the 1st Workshop on Context, Information and Ontologies*, page 3, 2009.

[12] M. Granitzer, A. S. Rath, M. Kröll, C. Seifert, D. Ipsmiller, D. Devaurs, N. Weber, and S. Lindstaedt. Machine learning based work task classification. *Journal of Digital Information Management*, 7(5):306–314, 2009.

[13] J. Hothi and W. Hall. An evaluation of adapted hypermedia techniques using static user modelling. In *Proceedings of the 2nd Workshop on Adaptive Hypertext and Hypermedia of the Hypertext*, volume 98, 1998.

[14] T. Joachims. Text categorization with support vector machines: Learning with many relevant features. *Machine Learning: ECML-98*, pages 137–142, 1998.

[15] M. Kellar and C. Watters. Using web browser interactions to predict task. In *Proceedings of the 15th WWW conference*, pages 843–844. ACM, 2006.

[16] B. Klimt and Y. Yang. The enron corpus: A new dataset for email classification research. *Machine Learning: ECML 2004*, pages 217–226, 2004.

[17] S. Koldijk, M. Sappelli, M. Neerincx, and W. Kraaij. Unobtrusive monitoring of knowledge workers for stress self-regulation. In *Proceedings of the 21th International Conference on User Modeling, Adaptation and Personalization*, 2013.

[18] S. Koldijk, M. van Staalduinen, M. Neerincx, and W. Kraaij. Real-time task recognition based on knowledge workers' computer activities. In *Proceedings of ECCE 2012 (Edinburgh, Scotland)*, 2012.

[19] M. Lansdale. The psychology of personal information management. *Applied Ergonomics*, 19(1):55–66, 1988.

[20] G. Manco, E. Masciari, M. Ruffolo, and A. Tagarelli. Towards an adaptive mail classifier. In *Proceedings of the Italian Association for Artificial Intelligence Workshop*, 2002.

[21] N. Oliver, G. Smith, C. Thakkar, and A. C. Surendran. SWISH: semantic analysis of window titles and switching history. In *Proceedings of the 11th IUI conference*, pages 194–201. ACM, 2006.

[22] J. Ruff. Information overload: Causes, symptoms and solutions. *Harvard Graduate School of Education*, pages 1–13, 2002.

[23] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz. A bayesian approach to filtering junk e-mail. In *Learning for Text Categorization: Papers from the 1998 workshop*, volume 62, pages 98–105, 1998.

[24] M. Sappelli, S. Verberne, and W. Kraaij. Using file system content to organize e-mail. In *Proceedings of the fourth symposium on Information interaction in context*, 2012.

[25] M. Sappelli, S. Verberne, and W. Kraaij. Combining textual and non-textual features for e-mail importance estimation. In *Submitted*, 2013.

[26] M. Sappelli, S. Verberne, M. van der Heijden, M. Hinne, and W. Kraaij. Collection and analysis of ground truth data for query intent. In *Proceedings of DIR 2012*, pages 7–10, 2012.

[27] A. Schick, L. Gordon, and S. Haka. Information overload: A temporal approach. *Accounting, Organizations and Society*, 15(3):199–220, 1990.

[28] J. Shen, L. Li, T. G. Dietterich, and J. L. Herlocker. A hybrid learning system for recognizing user tasks from desktop activities and email messages. In *Proceedings of the 11th international conference on Intelligent user interfaces*, pages 86–92. ACM, 2006.

[29] X. Shen, B. Tan, and C. Zhai. Implicit user modeling for personalized search. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 824–831. ACM, 2005.

[30] J. Spira and D. Goldes. Information overload: We have met the enemy and he is us. *Basex Inc*, 2007.

[31] K. Sugiyama, K. Hatano, and M. Yoshikawa. Adaptive web search based on user profile constructed without any effort from users. In *Proceedings of the 13th WWW conference*, pages 675–684. ACM, 2004.

[32] S. Verberne, M. Sappelli, and W. Kraaij. Term extraction for user profiling: evaluation by the user. In *Proceedings of the 21th International Conference on User Modeling, Adaptation and Personalization*, 2013.

[33] S. Verberne, M. Sappelli, D. Ransgaard Sørensen, and

W. Kraaij. Personalization in professional academic search. In *Workshop on Integrating IR technologies for Professional Search*, 2013.

[34] P. Warren, J. M. Gomez-Perez, C. Ruiz, J. Davies, I. Thurlow, and I. Dolinsek. Context as a tool for organizing and sharing knowledge. In *Proceedings of the 2nd International Workshop on Context, Information and Ontologies*.

APPENDIX

A. MOTIVATION FOR ATTENDING THE DOCTORAL CONSORTIUM

In this section, I will explain why I would like to attend the doctoral consortium at SIGIR.

I am currently in my second year of my PhD. The intended end-date is August 2015. My project is mostly on user modelling. However, we try to incorporate the viewpoints on user modelling from several research areas. We try to incorporate views from the area of cognitive psychology, views from human-computer interactions and the views from the field of information retrieval. This is a natural area for me to work in, since I have a background in (Cognitive) Artificial Intelligence. I specialized in Cognitive Research and Cognitive Engineering. Additionally, I obtained a bachelor degree in Linguistics with a focus on language technology.

Because of this variation it is sometimes difficult to find a good focus and to have a good overview of all the research that has already been done. Until now I have been focused mainly on text classification. I gained some experience with several term extraction techniques. There is a strong focus on the transparency of the textual features in my experiments (e.g. what can we deduce from the textual features and how well are the features interpretable for the user?).

Now, we shift our focus more to information retrieval. We aim to use information deduced from the user and his context to improve search results. Both my supervisors are experienced in the field of information retrieval. However, since they were my teachers during my studies, my knowledge on information retrieval is likely biased to their expertise. Since they are not specialized in user modelling within IR, I feel that the doctoral consortium at SIGIR may be a good platform to learn more on user modelling in IR. Hopefully this would give me inspiration for my research

Maya Sappelli
PhD Student
TNO & Radboud University Nijmegen

B. STATEMENT BY SUPERVISOR

Maya Sappelli will be halfway her thesis project at the time of SIGIR 2013 in Dublin. The project is scheduled to finish in the fall of 2015. This means that she has done quite a bit of literature study already and has conducted several experiments in different settings where user modelling or user interaction plays a role such as search intent classification and the contextual recommendation task at TREC. This experience has helped her to shape her research proposal, which is already in a good shape, but could be focused even more. The core of her work is to learn a faceted user model of interests, expertise, activities of a knowledge worker in an implicit fashion, just by observing activities such as reading writing, search by monitoring the interaction with a personal computing device. The idea is that such a faceted and time-annotated user model can enhance typical tasks such as internet search, local search, email prioritization and so on. The user model should be private and transparent, allowing for controlled sharing. The user model may be derived from a collection of unigram language models as its base form.

Maya would benefit from feedback in the areas of user models (elements, structure), minimal supervision of these models, design and management of longitudinal (e.g. several months) user studies with knowledge workers, how to keep them engaged etc. I think that the SIGIR doctoral consortium will be an excellent experience for Maya to discuss and learn about modelling and implicitly learning contexts (both from computational and information science and cognitive point of view) and evaluating these contexts in applications.

Wessel Kraaij, professor Information filtering and aggregation
Radboud University Nijmegen