Personalized Interaction with Semantic Information Portals

Eric Schwarzkopf
German Research Center for Artificial Intelligence (DFKI)
D-66123, Saarbruecken, Germany
Eric.Schwarzkopf@dfki.de

Abstract

We present the conception of a personalized interaction environment for semantic information portals, focusing on the user model maintained in this environment. The environment will support the user in finding documents via the portal, but also in managing and working with these documents once they have been retrieved. The user model is a central, explicit aspect of the user's interaction with the portal, with immediate access by the system as well as the user. This is made possible by the common semantic concepts available to both parties.

The described environment is currently in its early implementation phase. Our ideas and suggestions are presented here as points for discussion at the workshop.

1 Introduction

Accessing information via digital libraries and on-line portals has become more efficient thanks to advances in user interfaces as well as search and browsing technologies. Related research has focused largely on the improvement of the information retrieval process, but interaction with information does not end with retrieving relevant documents [Paepcke, 1996]: It also includes making sense of the retrieved information, organizing collected materials for near or long-term use, and sharing insights with colleagues.

With our current work, we aim to provide an interaction environment for information portals that facilitates these different kinds of interaction. The central component of this environment will be a personal document management system which will allow the user to organize and annotate collected information, thus facilitating the sense-making process and the expression of the user's gained knowledge. The user will be able to share the captured knowledge with colleagues and since this knowledge will be available to the information system, it will be utilized to improve the information retrieval process.

Since the user's knowledge is one of the most important factors in the interaction processes, finding an appropriate representation for this knowledge in the information system is essential for effective interaction. We decided to provide a common vocabulary (an *ontology*) as a basic framework for the personal *knowledge bases* to be created by the users. This allows us to model the user's knowledge on a more general level than would be possible when using only terms extracted from the information items the user collected. We expect that this representation will make it easier to communicate with the user (explaining decisions, requesting feedback) and utilize the targeted semantic information portals. In this paper, we will focus on the envisioned interaction with the user's personal knowledge base and the corresponding system-internal user representation.

The proposed interaction environment is currently in its early implementation phase, with only an early version of the document manager completed. Its conception, in particular that of the user model, is described here to get feedback and to provide points for discussion at the workshop.

In the following, we first present a short survey of related work we are building on. In section 3, we take a closer look at the proposed interaction environment. The conception of the user model is described in section 4, and we conclude by suggesting points for discussion.

2 Related Work

Information retrieval and interaction with (on-line) information resources is a well researched subject. We do not aim to reinvent the wheel, but to leverage existing technologies and experiences. In this section, we present only a small subset of the research we are building our work on.

2.1 Information Access and User Modeling

Mizzaro et al. [Mizzaro and Tasso, 2002] use personalization techniques to improve the effectiveness of information access systems. They focus on information retrieval and information filtering, utilizing a short-term and a long-term user profile, respectively. To create and maintain the long term model, they apply content-based techniques (natural-language processing, term co-occurrences), while the short-term model is made up from data about the user's current session (state of the system, actions performed by the user). The short-term model is used during the information retrieval process to provide terminological and strategic help, the latter being concerned with the strategy the user utilizes to organize the search process.

Billsus et al. [Billsus and Pazzani, 1999] use a user model consisting of a short- and a long-term profile for news filtering. The long term profile represents the general interests of the user and is represented by a naive Bayesian classifier (based on a set of pre-selected terms), while the short-term profile uses a nearest-neighbor approach to filter out news stories which are to similar to those already read by the user.

2.2 Information Management

VIKI [Marshall *et al.*, 1994] is a tool for spatially organizing electronic documents. VIKI focuses on the task of

making sense of collected documents. To this end, it allows the user to create an own "spatial language" by creating collections and composites out of the representations of documents.

NaviQue [Furnas and Rauch, 1998] is similar to VIKI, but integrates querying an external information source. For example, the user can navigate to a certain location in the information space, select a set of documents, and query the information source for similar items. Queries and the query-history are represented as objects in the UI and can be manipulated by the user.

2.3 Knowledge Representation and Capturing

Shum et al. [Shum *et al.*, 2002] attempt to capture scholarly argumentations by building up a network of argumentative claims. The relations and concepts this network is build from are defined in an ontology, and a web-based tool[1] allows researchers to build up their own claim networks or extend an existing one.

Trellis [Gil and Ratnakar, 2001] is a related tool which allows users to add their observations and opinions to online information resources via semantic annotations. The user starts with a certain claim, and uses the integrated query facility to find resources relevant to that claim. Using relations from a predefined vocabulary, the user expresses the relationships between resources and how they support or refute the claim. A demo version of Trellis is available on-line.[2]

3 Interaction Environment

With the envisioned interaction environment, we aim to cover a larger part of the different interactions with information than only the information retrieval process. To this end, we plan to provide a personal document manager (see figure 1), which will assist the user in making sense of collected information, organizing it for short- and long-term use, and sharing gained knowledge. It will be interoperating with a recommendation system which sits on top of an information portal (forming its knowledge base) and utilizes the additional data available through the document manager to provide personalized support to the user.

The following scenario provides one example of the envisioned interaction between the user and the information system.

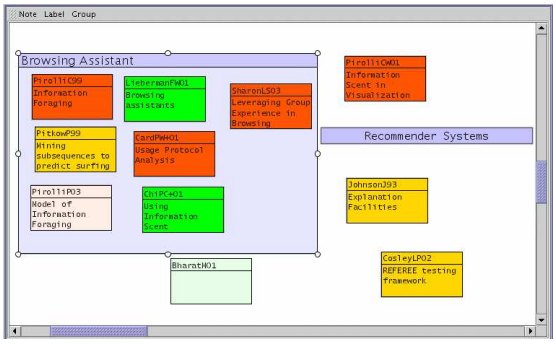

Figure 1: The document manager allows users to place objects freely on a 2-dimensional *whiteboard*

Imagine you are collecting literature relevant to your current research project. Using the query functionality of the information portal, you find an initial, small set of related documents and place them via drag-and-drop into your personal document collection. You create a label on the whiteboard representing your project, place the documents close to that label, and take a closer look at them.

One of the documents includes a nice example of your project's main issue, so you define a `provides-example-for` relation between the document and your project by dragging the document representation onto the label representing your project and selecting the corresponding relation in the appearing pop-up menu. In addition, the document provides a good theoretical foundation for further work, so you mark it as `theoretical` work and `important` by selecting the corresponding options from the pop-up menu associated with the document representation. Finally, you decide to associate a note with the document pointing out the most important claims made in it.

Another document provides the description of an algorithm for a more specific sub-issue of your project, and a corresponding empirical evaluation. So, you decide to create a new group on the whiteboard representing the sub-issue, define an `addresses` relation between the project and the issue, and add the document to the group. Then, you mark the document as containing `theoretical` and `empirical` work.

After reading and annotating the remaining documents, you decide to look for up-to-date work related to the sub-issue you identified earlier. To this end, you select the group representing the issue and the appropriate recommendation need from a pop-up menu. As a result, the document manager connects you with the information portal and automatically requests a list of candidate documents, which is displayed in a web-browser window. In a second window, you see with which general topics known to the information portal the system associates your issue (with an indication of the confidence these system has in these relations). You are not really satisfied with the system's classification, so you decide to deactivate context-aware query support and manually provide a set of query terms. The resulting set of documents seems to contain material somewhat related to your problem, and the associated topics seem to fit better than the ones proposed by the system. So, using the window showing the system's initial classification, you remove the proposed topics and manually add the appropriate topics from the list of query results. Then, you reactivate context-aware support which causes the system to evaluate the documents in the result set in respect to the topics and documents associated with the issue you are interested in.

Because the system still seems to be somewhat off track, you decide to ignore its evaluations for now and continue browsing. One document in the result set leads you to the proceedings of a conference which contain at least two very interesting publications. You add these documents to your personal collection, placing them in the group representing the issue of interest. As a result of adding this new data, the system informs you about three papers which are more current and, with high confidence, relevant to your current work. You take a closer look at the recommended documents and they are indeed very valuable to you.

After adding the recommended documents to your collection, you change your goal to finding the most relevant work in general related to your project. This time, thanks to the amount of data you have already collected, the system

[1] http://kmi.open.ac.uk/projects/scholonto/

[2] http://www.isi.edu/ikcap/trellis/

provides an adequate classification of your project, and the proposed documents look very promising.

There are some important things to note about this scenario:

- The envisioned interaction components provide an additional layer of support on top of an existing information retrieval system.
- The components take the user's current working context and goals into account.
- The user always has the possibility to check how the system interprets the user's current working context and interests.
- The user can explicitly modify this interpretation.
- The user can always deactivate or reactivate contextualized interaction.

Central to the presented interaction is the system's model of the user, which is described in more detail in the next section.[3]

4 User Modeling

The information system maintains a user model to approximate the user's information need. Applications of this user model include

- Automatic Query Expansion: Given a set of query terms supplied by the user, identify additional terms to increase the accuracy of the query.
- Result Filtering: Given a set of documents (possibly the results of a query), evaluate their relevance.
- Recommendations: Given a recommendation need, automatically select the most appropriate information items and suggest them to the user.

4.1 Requirements

For the envisioned interactions, the user model has to fulfill several requirements:

- It has to make sense to the user, so it can be inspected and, if necessary, corrected.
- The user will maintain a set of working contexts (e.g., materials related to a specific project). The user model has to take the currently active context into account.
- It has to be comparable to other user models in order to be usable by a collaborative filtering recommender.
- It has to be mappable into the ontology of the information portal, so it can be utilized by a knowledge-based recommender.
- It has to be mappable into document space, so it can be utilized by a content-based recommender.
- It has to be as accurate as possible, in order for the recommendation components to provide accurate recommendations and evaluations.
- It has to be quickly adaptable to the user's changing information need.

[3]Note that the knowledge base of the information portal makes it possible to make the user model an explicit part of the interaction, instead of being implicitly created and updated by the system on the basis of query terms and relevance feedback.

Some of these requirements contradict one another: A user model that adapts quickly on the basis of uncertain evidences will not be very accurate, and a model that can be mapped into document space is probably not very intelligible to the user.

4.2 Conception

To fulfill the requirements listed above, we propose a user model which consists of three different layers:

- A general, fairly accurate, only slowly adapting long-term profile. This layer represents the user's general interests in terms of research areas. For example, the long-term profile might state that the user is quite knowledgeable in "Machine Learning" and "User Modeling", but less so in "Computer Graphics".
- A more specific medium-term profile which represents the user's current working context. It consists of the most specific research areas covering the documents in the context, concepts associated with these documents (such as authors and conferences), as well as a representation of the documents themselves. For example, a context of a user working in uncertain reasoning might contain references to the research areas "Bayesian Networks" and "Markov Processes", the "Uncertainty in AI" conferences, and document-terms like "reasoning", "temporal models", "memory", and "complexity".
- A very specific, quickly adapting, but possibly less accurate short-term profile. It represents the information need during the user's current interaction with the system, and contains the same kind of information as the medium-term profile, but puts a stronger emphasis on document terms than on more general ontological concepts. The main difference to the medium term model is that the short-term model is adapted after each interaction with the portal, which means it has to cope with few and uncertain evidences, as well as hard time constraints for analyzing them.

Having a model consisting of three layers leads to the issue of how to combine these layers. The long-term model is fairly accurate, but too general to be of much use on its own. On the other hand, the short-term model is the most relevant for the current interaction, but it can be quite inaccurate. One possible technique for combining the profiles is to weight each layer according to the confidence the system has in its appropriateness, and then use a weighted sum over the layers to create final weights for the profile's contents. Thus, when the user has just started browsing though the portal, the confidence in the short-term profile will be low and the recommender will rely on the medium and long-term profiles. But after collecting additional evidences, the confidence in the short-term profile will increase and it will gain a greater influence on the overall behavior of the system.

With this rather abstract view of the user model established, the next issue is how to actually realize it and and make it controllable and transparent to the user. Here, the basic idea is to exploit the personal knowledge base the user creates with the help of the document manager.

This knowledge base contains the documents, projects, and issues, as well as the relations between these objects important to the user. Thus, it approximates the user's interests and state of knowledge. If we assume a well maintained personal knowledge base is available, the long-term

profile can be generated by weighting the importance of objects in this knowledge base (for example, by taking the connectedness of objects into account, or exploiting techniques from the area of social network analysis [Wasserman and Faust, 1994]), and utilizing the ontology to identify the high-level research areas associated with the most highly ranked objects.

For the medium-term model, the user's interaction history with the knowledge base should be taken into account. Objects could receive an importance value according to the kind of interactions, their frequency of occurrence, and how long ago these interactions took place. From these *hot spots*, a value representing importance could be spread a short distance throughout the knowledge base to adapt the importance values of closely related objects.

The short-term model is a special case, because it takes the interaction with the information portal into account and thus has to use objects which are not part of the personal knowledge base. Here, the principal idea is to map the relevant part of the user's personal knowledge base into the knowledge base of the information portal, and then exploit the data about the user's interaction with the portal to distribute importance values similar to how it is done for the medium-term model.

Exploiting the personal knowledge base is only possible and reasonable if it is created and well maintained by the user. But the user should not be forced into using (nor expected to use) a predefined, rigid model and update formalism, especially not in the application domain of the document manager. Thus, we expect having to cope with rather pragmatically constructed knowledge bases. To lessen the impact of this on the quality of the user model, the document manager will try to automatically extend the knowledge base by adding relations between objects if these are not defined by the user. For example, if the user adds a new object to the collection, but does not define any relations, the system will attempt to assign a topic and add relations based on examples derived from object and relations added earlier to the knowledge base. The automatically added relations are distinguished from those defined by the user: There is a lower confidence in their correctness and so they do not have as much influence in later computations as the user defined ones.

Now, what about transparency and controllability? By using the personal knowledge base as the base of the user model and its basic representation, the user actually interacts with the user model when interacting with the document collection. That is, there is no "external" user model the user has to care about. Since all objects (including relations) in the knowledge base have an explicit representation in the document manager, the importance assigned to an object, and the objects contributing to that importance, could be visualized by coloring objects and relations similar to a thermal image. Thus, objects of high importance would be marked as being "hot" and, correspondingly, objects of low importance as being "cold". By being able to change the temperature of an object, the user would have an immediate way to influence the current representation of the user's information need.

5 Contributions and Issues for Discussion

We presented an overview of the conception of an interaction environment for semantic information portals, focusing on issues related to the user model. The described system is currently in its early implementation phase. We suggest the following issues for further discussion at the workshop:

- Is the suggested user model appropriate for the envisioned interaction environment?
- With the user model as a central, explicit part of the interaction with the information system, in what ways should the user be able to modify and adapt it (e.g., reducing the model to a subsection of the knowledge base for the current interaction, or loading older profiles into the system)?

6 Acknowledgments

I thank Anthony Jameson for his advice and ideas during the ongoing development of the project.

The described work is being funded by the German Ministry of Education, Research, Science, and Technology under grant 08C5940 as part of the SemIPort project.

References

[Billsus and Pazzani, 1999] Daniel Billsus and Michael J. Pazzani. A Personal News Agent that Talks, Learns and Explains. In *Proceedings of the Third International Conference on Autonomous Agents*, pages 268–275. ACM Press, 1999.

[Furnas and Rauch, 1998] George W. Furnas and Samuel J. Rauch. Considerations for Information Environments and the NaviQue Workspace. In *Proceedings of the Third ACM Conference on Digital Libraries*, pages 79–88. ACM Press, 1998.

[Gil and Ratnakar, 2001] Yolanda Gil and Varun Ratnakar. TRELLIS: An Interactive Tool for Capturing Information Analysis and Decision Making. Technical report, The University of Southern California/Information Science Institute, California, USA, 8 2001.

[Marshall *et al.*, 1994] Catherine C. Marshall, Frank M. Shipman III, and James H. Coombs. VIKI: Spatial Hypertext Supporting Emergent Structure. In *Proceedings of the ACM European Conference on Hypertext (ECHT'94)*, pages 13–23. ACM Press, 1994.

[Mizzaro and Tasso, 2002] Stefano Mizzaro and Carlo Tasso. Ephemeral and Persistent Personalization in Adaptive Information Access to Scholary Publications on the Web. In Paul De Bra, Peter Brusilovsky, and Ricardo Conejo, editors, *Adaptive Hypermedia and Adaptive Web-Based Systems, Second International Conference AH2002*, pages 306–316. Springer, 2002.

[Paepcke, 1996] Andreas Paepcke. Digital Libraries: Searching Is Not Enough. *D-Lib Magazine*, 2(5), 1996.

[Shum *et al.*, 2002] Simon Buckingham Shum, Victoria Uren Gangmin Li, John Domingue, and Enrico Motta. Visualizing Internetworked Argumentation. In Paul A. Kirschner, Simon J. Buckingham Shum, and Chad S. Carr, editors, *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, pages 185–204. Springer, 2002. In Press.

[Wasserman and Faust, 1994] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.