# From User Query to User Model and Back: Adaptive Relevance-Based Visualization for Information Foraging

Jae-wook Ahn and Peter Brusilovsky
*School of Information Sciences, University of Pittsburgh*
*{jaa38, peterb}@pitt.edu*

## Abstract

*Adaptive information filtering is a promising tool for both casual Web news readers and professional intelligence analysts. Adaptive filtering augments the traditional query- or profile-based rankings provided by search engines. An interesting research challenge in this context is to offer users more control over the rankings by letting them mediate between the two extremes – query- and profile-based rankings. To address this challenge, we developed an adaptive relevance-based visual exploration tool based on the VIBE (Visual Information Browsing Environment) visualization approach, which was previously developed at our School. This paper presents the rationale and functionality of this visual exploration tool and reports the results of its preliminary evaluation.*

## 1. Introduction

Adaptive information filtering [3] is emerging as a popular Web information access technology. An adaptive filtering engine collects potentially useful information about the user's interests, preferences, and knowledge, either from explicit feedback, originating from the user's relevance judgments, or implicit feedback, derived by observing the user's search and browsing activities. The engine then uses such information to predict and recommend information that is potentially relevant to the user. This provides a useful alternative to traditional "one-size-fits-all" search engines, which respond to a user's query with a list of links to information resources ranked by their relevance to the query. In several relevant contexts, such as news reading [2] or TV program selection [7], personalized information filtering tools have already demonstrated their value.

In the context of the DARPA GALE (http://www.darpa.mil/ipto/ programs/gale) project, we explore the potential use of adaptive information filtering in the area of intelligence analysis, where the use of personalized tools is still very rare [11]. We think that adaptive filtering can augment the traditional search-based information access at the information foraging stage of the analysts' work [9], when they are attempting to collect potential useful information from documents in various media and sources.


**Figure 1. ROSETTA System Interface**

The ROSETTA system (Figure 1), developed by our joint project team as an information analyst's "co-pilot," includes both search and personalized filtering. The latter is supported by the personalized filtering engine CAFÉ (Carnegie Mellon Adaptive Filtering Engine), which collects various kinds of implicit user feedback and builds a profile of user interests. CAFÉ processes information retrieved by the search component of ROSETTA and re-ranks it according to the user profile. When Rosetta was extensively evaluated in two rounds, by professional intelligence analysts, CAFÉ was highly praised. However, users also indicated that they wished they had more control over the performance of the engine. In particular, the users were interested in having more control over the ranking by being able to mediate between the two currently available extremes – query-based ranking in

the search component and profile-based ranking in the filtering component.

The problem of "fusing" query- ranking and profile-based rankings is not new. The traditional solution to this problem, which is applied in several adaptive search systems [6], is to select a fixed mediation point α between 0 and 1 and to produce a personalized rank by fusing query- and profile-based rankings with coefficients α and (1-α). By manipulating α, the system designers can give more priority to documents similar to the query or documents similar to the profile. However, this solution is not providing the analysts with the flexibility they desire. This paper presents a more flexible approach to "fusing" query- and profile-based rankings. The idea of this approach is to allow the analysts to dynamically decide whether they are interested in documents which are closer to the query or documents which are closer to the user profile – with the ability to navigate on a continuum between the query to the user profile and back again.

The core component of our approach is the relevance-based visualization originally implemented in VIBE [8]. VIBE is known as an excellent tool for visual query results analysis. It allows the user to explore the connection between search results and query terms, for example, enabling the user to pick a subset of results that is more relevant to a specific query term or group of terms. Our visual analysis tool applies relevance-based visualization to help the user to mediate between the query terms and terms from the user profile. The application of the user profile makes the relevance-based visualization adaptive. The results of the visualization are different for different users who have submitted the same query and even different over time for the same user, if the interests of the user represented in the user profile evolve. This paper presents our implementation of VIBE for adaptive relevance-based visualization, stresses several features that are critical for this type of visualization, and presents some evaluation results.

## 2. Fusing Query- and Profile-based Rankings with VIBE

VIBE was originally developed at Molde College in Norway and the School of Information Sciences at the University of Pittsburgh [8]. It is a document visualization tool, which supports POI (Point of Interest)-based browsing. POIs represent key concepts or keywords and are displayed as user-draggable icons on the screen. The documents are placed according to their similarities to the POIs. The main idea is that if a document is more similar to POI $P_a$ than POI $P_b$, then it is placed closer to $P_a$ than $P_b$ and the closeness is determined by the document-POI similarity ratio. For example, if a document has similarity 0.3 and 0.6 to POI $P_a$ and $P_b$ respectively, the similarity ratio to these two POIs is 1:2 and the document is placed at a one-third position from $P_b$ on the line connecting those two POIs, because it is twice as similar to $P_b$ than $P_a$. Detailed algorithm for placing a document among multiple POIs was presented in [8]. Users can drag and move POIs anywhere they want and the locations of the documents are dynamically updated depending on their similarities to the POIs. They can easily find out which documents are more similar to a certain POI by their locations and they can also find out the degree of similarity by the documents' degree of movements (if documents follow the movement of a POI a lot, then they can be thought of as very similar to the POI). VIBE is a relatively popular visualization approach. Several implementations of VIBE and similar systems inspired by VIBE were used for the relevance-based analysis of query results, with query terms used as POIs.

As we mentioned above, our key idea for using VIBE as a query to profile mediation tool is to use both query terms and profile terms as POIs. To develop this kind of visualization we started with our own applet-based version of VIBE, which implemented basic functionalities [8], and equipped it with various new features to support efficient query-to-profile mediation. These features will be explained in detail later.

### 2.1. Ranked Lists vs. VIBE-based Fusing

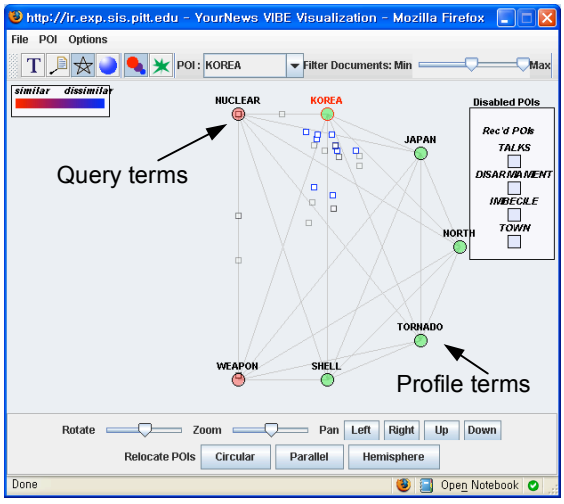


**Figure 2. VIBE Visualization for Query and Profile-based Ranking Fusion**

Figure 2 shows an example of applying VIBE to the query and profile fusion problem. First and foremost, we can see 7 circles colored in pink and green. These

are POIs representing two different sets of terms: query terms (pink) and profile terms (green). In this example a user entered a query "NUCLEAR WEAPON" and the system retrieved relevant articles with high similarity scores (which will be discussed in detail in Section 3). White squares represent these retrieved documents and users can examine their titles and summaries by hovering the mouse cursor over the square icons. We extracted 10 profile terms and displayed the top 5 of them as green circles on the screen. The rest of the profile terms are disabled temporarily and docked in a white box at the corner of the screen (4 in this case because one term, NUCLEAR, overlapped the query). Users are able to freely move both query and profile terms and explore which document is related to which POI (or term).

This example clearly demonstrates the difference between the traditional search result (query-based ranked list), an adaptively re-ranked result (profile-based ranked list) and our flexible approach exploiting VIBE. Originally, the search engine results contain the top 5 articles on Iranian nuclear weapon development. The ranked list sorts the documents by their relevance score and users typically examine the top ones first. This result is appropriate if the user in this example was most interested in recent events in Iran. However, let's consider a user who is interested in Korean affairs including North Korean nuclear weapon development. Over the weeks of using the system, this user has been accumulating terms like KOREA, NUCLEAR, JAPAN, and NORTH in her profile of interests (revealed in Figure 2 as green circles). Proponents of adaptive search and filtering systems would argue that this user would be most interested to see information about North Korean, not Iranian nuclear programs and would prefer to see news ranked according to her profile with North Korean news emerging on the top of the list. Unfortunately, in a realistic context it is hard to decide what is the real need of the user because of a lack of information. Her interests may have remained the same (i.e., she does prefer news on North Korean nuclear developments) or may have switched to a different direction (i.e., she is interested in seeing up-to-date news about other programs). Of course, the user could have entered a query like "NORTH KOREAN NUCLEAR WEAPON" more explicitly but it is a very well-known phenomenon that most users are not familiar with expressing their needs in exact query terms and the number of terms used for their queries are just two or three in average [5].

"Fusing" query-based ranking and profile-based ranking is a more reliable way to assist the user in an ambiguous context. As we pointed out, traditional fusing approaches [6] attempt to create a mixed ranked list, which lacks flexibility and can be confusing to the user since the ranking order can be hard to explain. In contrast, we provide a visualization using VIBE and let users interactively explore the query terms, profile terms, and the retrieved documents simultaneously. The users are able to understand the relationships among these three components and discover relevant information more easily. The example above clearly shows the benefits of our approach. By examining the locations of the articles using VIBE, it is surprising that a lot of articles are placed closer to a profile term (KOREA) than the query terms (NUCLEAR and WEAPON). This result is very interesting because the documents visualized here are exactly the same set of articles displayed in the query-based ranked list retrieved by a conventional search engine, where the top 5, most important articles were about Iranian nuclear weapon development. Our approach can provide users with the flexibility to intuitively discern which documents are more related to the query or the profile terms by just glancing at the picture. We don't have to choose either of the two: query or user profile-based ranked list as in [6]. We can merely show the relatedness of each document to each of the concepts and let users visually explore to understand what the situation really is.

In terms of the example here, there are two cases: (1) the user is interested in the nuclear weapon in general, or (2) she is more interested in Korean nuclear weapon development. In the first case, the system is able to provide the user with more specific information about the search results (which article is more about a specific sub-category like Iranian or North Korean nuclear weapon development) as well as serving the original user needs. For the second case, the user can use a visual aid to easily locate interesting articles, which would have been rather hidden in a conventional, query-based ranking. In the following sections, we will introduce more detailed descriptions about the components of the VIBE framework.

## 2.2. VIBE Tools

Our implementation of VIBE includes new as well as original tools to help users more efficiently understand the relationships among POI and documents. Several of these are worth being discussed in the context of query-to-profile mediation.

*Document trail* – one of the basic features VIBE supports, document trails are a simple but powerful tool to explain the document-POI relationship. When a user drags and moves a POI, then its related (similar) documents follow the movements of the POI. The trails can reflect the movements of multiple documents and they remain until the next update of the screen, so that users can easily compare the length of the trails to

know which ones are longest, meaning they are the most similar ones to the corresponding POI (query or profile terms in the context of this study).

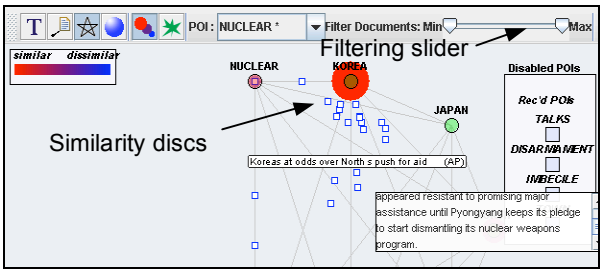


**Figure 3. Similarity Discs and Document Filtering Tool**

*Document similarity disc* – As in Figure 3, users hover a mouse cursor over a document and then discs with various size and colors are displayed on each POI (query or profile terms). The size and color (red to blue spectrum) of the discs here represent the similarity between the corresponding document and the POIs: bigger size and red color means higher similarity. Or inversely, they can place the mouse cursor over a POI, then the discs are overlaid on documents this time. Therefore, users are able to locate similar documents to a POI and understand how closely they are related to the POI by just glancing at the size and color of the discs.

*Document filtering* and *Distortion tools* – users can feel that they need to filter out some of the documents from the screen. A document filtering function can meet this need (Figure 3). Users can select a POI by clicking on it or from a drop down menu, and then set a low and high threshold to filter out documents outside of this range interactively using a double slider. Our implementation of VIBE also supports some standard distortion tools: panning, zooming, and rotating of view-ports. These are easily manipulated by simple mouse or keyboard actions.

## 2.3. POI Allocation Presets

The original VIBE provided an initial allocation of POIs as a circle and the documents are placed inside it (Figure 4 above). In order to better support the fusion of query and profile-based ranked list visualization, we added two more POI allocation schemes: Hemisphere (Figure 2) and Parallel (Figure 4 below). In the Hemisphere preset, which was already introduced above, the original circle is divided into left and right parts and the query terms are placed only in the left hemisphere and the profile terms are placed in the right hemisphere. Thus, a user can compare the different distribution of documents ruled by two types of POIs with more ease. Another preset is Parallel, which allocates query terms on one vertical line and profile terms on another vertical line. Two lines of profile terms are located at the left and rightmost edges of the screen, so that users should discover which side the retrieved documents are closer to: query terms or profile terms. If we look at Figure 4 (below), we can notice some documents are on the left vertical line which connects two query terms. It means those documents are totally related to this query only and not related to the user profile at all. However, we can discover many more documents in the middle of the two vertical lines, which suggests that they are related to both query and profile terms and because their position is a little bit closer to profile terms, they should be more about the profile terms KOREA, JAPAN, and NORTH. TORNADO and SHELL are irrelevant to the retrieved documents and can be ignored, because the documents are very far from them.

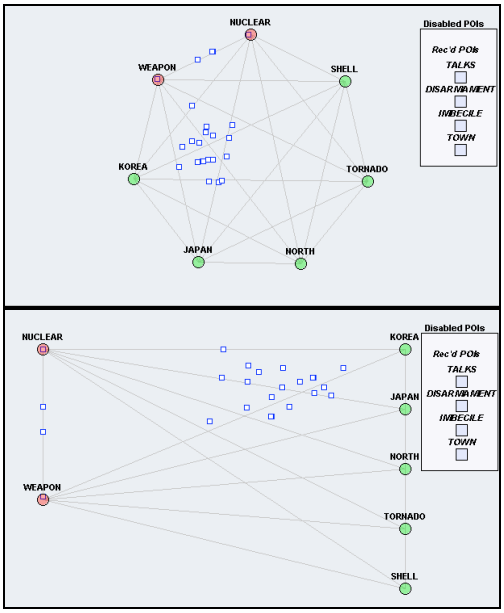


**Figure 4. POI Presets – Parallel and Circular**

## 2.4. POI Dock

In general, it is better not to have too many POIs displayed on the screen at the same time in order to avoid high cognitive load and ambiguity. It has never been a problem for the traditional use of VIBE since the number of terms in a query is typically sufficiently small. However, a user profile may contain dozens of reasonably weighted terms. Our analysis of typical user profiles demonstrated that for the purpose of query-to-profile mediation, VIBE should allow the analysts to explore between 10 and 15 profile terms. To support

this, we developed a new feature that we call POI dock. POI dock allows the user to enable/disable POIs temporarily. Disabled POIs are docked in a small box at the corner of the screen and they can be dragged out to be enabled anytime again. In our ranked list fusion system, profile terms with less importance are treated as initially disabled POIs and displayed in the POI dock (white boxes in Figure 2, 3, and 4). However, it's up to the user to decide which of the profile terms should be used for exploration of the current query and which not. Terms that have high weight in the profile, but are not relevant to explore in the context of a query could be docked and disabled. Vice versa, some weaker profile terms that were originally docked, are now considered as interesting to explore, and can be moved out from the dock and enabled. Our own experience with the system showed that the POI dock is really important for the task of query-to-profile mediation and the study reported below confirmed it.

## 3. YourNews: The Evaluation Context

As a production-oriented system, ROSETTA is too large and has too many features to serve as a platform for a focused study. To evaluate new interface and personalization features for ROSETTA, we developed a compact personalized news access system, YourNews, which can be easily extended with different innovative features and serve as a platform for their evaluation. Among the innovative features evaluated with YourNews in the past was the application of open user profiles to improve information filtering [1]. We also selected YourNews as a platform for our evaluation of query-profile mediation with VIBE.

YourNews gathers news items from a range of RSS news feeds and organizes it into eight (8) topics, which are shown to the user as tabs (Figure 5). The user can then search for news items and receive system recommendations within each of the eight topic tabs. To support news search and filtering, the system uses the well-known vector space model [10]. The system represents news items as document term vectors. The terms are stemmed according to the Krovetz algorithm [4] and weighted using TF-IDF.

YourNews provides adaptive news filtering by tracking news articles read by a user, building a profile of user interests, and recommending articles that match user interests. The interest profile is generated by creating a vector of weighted terms, found in the user's reading choices. For each user, the system maintains 16 interest profiles, which correspond to eight supported news topics which each have two different profile types: long and short term. Short term profiles, representing the user's current interests, are extracted from their 20 most-recently read news articles, while long term profiles represent general user interests. The actual recommendation process is achieved by calculating similarity scores between the profile vectors and the news article vectors. The articles with higher scores are recommended to users.

The presence of query-based search and profile-based filtering makes YourNews the perfect platform to evaluate VIBE visualization, which fuses query- and profile-based ranked lists. For the study, the VIBE applet was connected to YourNews search interface by a link. After users entered query terms and received search results (Figure 5), they click on this link to open VIBE applet window where VIBE presents search results (news articles). The query terms and top user profile terms are passed to VIBE as POIs, so that the user could explore the search results by mediating between these POIs.

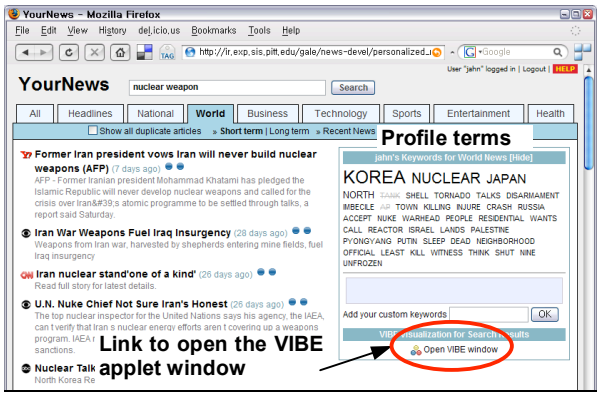


**Figure 5. YourNews System**

## 4. Evaluation

In order to assess the effectiveness of query-to-profile mediation with VIBE, we conducted a pilot study. For the study we recruited seven subjects who participated in our previous study with the YourNews system [1]. All subjects were graduate Information Science students – the most similar to the intelligence analyst type of subjects which we could find in a university. The subjects had extensive experience with the YourNews system before it was extended with search and VIBE functionalities. After watching a brief demonstration of VIBE and its main features, the subjects were given time to freely explore new features of YourNews (search and VIBE). After that, they were asked to answer several questions about their experience.

The questions attempted to solicit user feedback of both new features – search and VIBE. While the main goal of our evaluation is VIBE, the pre-condition of

using VIBE is that the system already possesses good search and filtering components. The search and filtering components of ROSETTA were already evaluated in earlier studies as well as in the filtering component of YourNews [1]. However, YourNews search has not been evaluated in the past. Thus, along with the evaluation of VIBE, we had to assess the performance of our search engine, which passes the retrieved documents to the VIBE engine. The questions were constructed to prove the following hypotheses.

(1) The new search function could provide good search ability
(2) The VIBE visualization, used for the fusion of query-based and profile-based rankings, was able to satisfy the subjects.

### 4.1. Subjective Feedback on Baseline Search

In order to find out the subjects' satisfaction level on the baseline search function, we broke down the big hypothesis and asked 5 questions covering the following aspects.

(1) Whether the system returned sufficient search results
(2) Whether the retrieval results were precise
(3) Whether the subjects were confident with the results
(4) Whether the subjects felt it was easy to formulate queries
(5) Whether the subjects had positive experience with searching

The subjects were asked to respond on five-point Likert scale (from 1=Not at all to 5=Extremely). Their answers are summarized in Figure 6. From the results, we can quickly observe the subjects were positive about the system. More than 70% of the subjects answered positive (score 4 or 5) than neutral (3) or negative (1 or 2) for all 5 questions. Interesting aspects concern precision and confidence, where more than 80% of the subjects answered positive or strongly positive but negative responses were also discovered. It can be understood as a higher variance and difference among users who use search systems, which confirms again the need for more sophisticated solutions other than conventional searching for plain users (such as the personalized search or the proposed approach in this study).

### 4.2. Subjective Feedback on VIBE Visualization

We asked 6 questions as in the previous survey and used the five-point Likert scale to collect user satisfaction information about the VIBE visualization for query and profile-based ranked list fusion.

(1) Whether it was easy to interpret the visualization
(2) Whether it was easy to work with POIs
(3) Whether the subjects found the fusion approach useful
(4) Whether the subjects found the multiple POI presets useful
(5) Whether the subjects found the POI dock useful
(6) Whether the subjects found the similarity disc useful

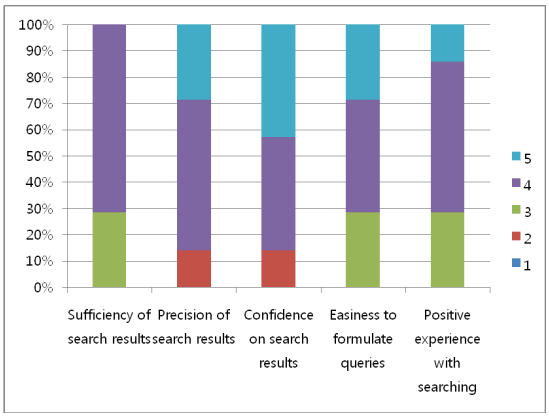


**Figure 6. User Feedback on Search Function**

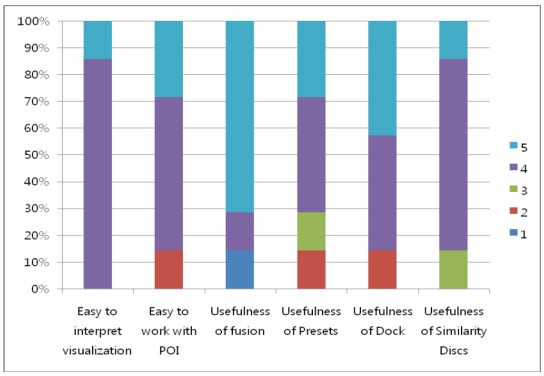


**Figure 7. User Feedback on VIBE Visualization for Fusion**

The results are provided in Figure 7. This time the subjects showed a slightly higher satisfaction level than the baseline search function. On 5 out of 6 questions, more than 80% of the subjects responded positively and on the remaining one question, 70% of the subjects returned positive answers. It was very encouraging that all subjects found the visualization results easy to interpret. However, about the main focus of this study, "Usefulness of fusion," we were able to observe one extreme value "It was not useful at all" given by one single subject. It's interesting that all other subjects gave very high scores, 4 or 5 – not even neutral – and only this subject showed a sudden drop (scored as 1) on this question. This subject also answered negatively about the ease of work with POI,

which suggests that she might not have been very successful working with the basic functionality of the system. This result can be anticipated because the VIBE framework itself is easy to interpret (from the answer to the first question) but manipulating POIs which represent the fusion of heterogeneous information would require a higher cognitive load for some users. We can also consider a possibility that some users can find visual tools less useful than in general. In some instances, a correlation between the users' attitude toward visualization tools and their spatial capability were found, such that they depend less on visual information and thus finds visualization less useful than other people. Even though we were not able to test each subject's spatial capability, we are going to investigate this aspect in our future study which will be more elaborate and contain an eye-tracker.

We also asked the subjects to pick the most useful feature introduced in this study. 4 out of 7 subjects chose the ranked list fusion function as their most useful feature, 2 subject picked POI dock, and 1 liked the Similarity disc tool. This result confirms the previous result, where about 85% of the subjects agreed to the usefulness of the fusion.

## 5. Conclusions and Future Work

This paper proposes adaptive relevance-based visualization as a tool to help intelligence analysts in the process of information foraging. The proposed visualization approach allows the analysts to mediate between a query and an accumulated interest profile. We developed a specific version of the VIBE visualization system, which was extended with several tools to assist query-to-profile mediation. The system was explored in the context of a personalized news system, YourNews. The subjects in our study positively evaluated the relevance-based visualization approach and supporting exploration tools. The results encourage us to proceed with installation of the developed system in ROSETTA, an analyst's workbench developed by our joint project team.

We plan to continue the exploration of the relevance-based visualization approach in the context of intelligence analysis. Among our specific goals is an eye-tracking study of VIBE and an analysis of individual differences and their correlation with user productivity and satisfaction with VIBE. In our earlier studies we collected some evidence that compared to average users, the users with low spatial abilities do not benefit from visualization and have a lower opinion of it. We want to check whether the observed difference in user satisfaction with VIBE may be caused by specific individual traits.

## 6. Acknowledgements

This paper is partially supported by DARPA GALE project and by the National Science Foundation under Grant No. 0447083.

## 7. References

[1] Ahn, J., Brusilovsky, P., Grady, J., He, D., and Syn, S. Open User Profile for Adaptive News System: Help or Harm? In: *Proceedings of 16th International World Wide Web Conference*, Banff, Alberta, CANADA, May 2007.

[2] Billsus, D. and Pazzani, M. J. User modeling for adaptive news access. *User Modeling and User Adapted Interaction* 10 (2-3), 147-180, 2000.

[3] Hanani, U., Shapira, B., and Shoval, P. Information filtering: Overview of issues, research and systems. *User Modeling and User Adapted Interaction* 11 (3), 203-259, 2001.

[4] Krovetz, R. Viewing Morphology as an Inference Process. In: *Proc. of ACM SIGIR'93*, 191-203, 1993.

[5] Lawrence S. and Giles C.L.: Context and Page Analysis for Improved Web Search. IEEE Internet Computing, 38–46, July-August 1998.

[6] Micarelli, A., Gasparetti, F., Sciarrone, F., and Gauch, S. Personalized search on the World Wide Web. In: P. Brusilovsky, A. Kobsa and W. Neidl (eds.): *The Adaptive Web: Methods and Strategies of Web Personalization*. Lecture Notes in Computer Science, Vol. 4321, Berlin Heidelberg New York: Springer-Verlag, pp. 195-230, 2007.

[7] O'Sullivan, D., Smyth, B., Wilson, D. C., McDonald, K., and Smeaton, A. F. Improving the Quality of the Personalized Electronic Program Guide. *User Modeling and User Adapted Interaction* 14 (1), 5-36, 2004.

[8] Olsen, K. A., Korfhage, R. R., Sochats, K. M., Spring, M. B., and Williams, J. G. Visualisation of a document collection: The VIBE system. *Information Processing and Management* 29 (1), 1993.

[9] Pirolli, P. and Card, S. K. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In: *Proceedings of International Conference on Intelligence Analysis*, 2005.

[10] Salton, G. (ed.) The SMART retrieval system: experiments in automatic document processing. Prentice-Hall, Englewood Cliffs, NJ, 1971.

[11] Santos Jr., E., Nguyen, H., Zhao, Q., and Wang, H. User modeling for intent prediction in information analysis. In: Proceedings of 47th Annual Meeting for the Human Factors and Ergonomics Society (HFES-03), Denver, CO, pp. 1034-1038, 2003.