Envisioning User Models for Adaptive Visualization

1st Author
1st author's affiliation
1st line of address
2nd line of address
Telephone number, incl. country code
1st author's email address

2nd Author
2nd author's affiliation
1st line of address
2nd line of address
Telephone number, incl. country code
2nd E-mail

3rd Author
3rd author's affiliation
1st line of address
2nd line of address
Telephone number, incl. country code
3rd E-mail

ABSTRACT

Adaptive search systems apply user models to provide better separation of relevant and non-relevant documents in a list of results. This paper presents our attempt to leverage this ability of user models in the context of visual information exploration. We developed an adaptive visualization approach for presentation and exploration of search results. We simulated a visual intelligence search/analysis scenario with log data extracted from an adaptive information foraging study and were able to verify that our method could improve the ability of traditional relevance visualization to separate relevant and irrelevant information.

Keywords

Visualization, VIBE, User model, Query, Personalized Search.

1. INTRODUCTION

User modeling is a technique to support adaptive information access by supplementing non-adaptive information access systems with information about its users. Over the last 20 years, user models have become very popular in the area of information access systems [1]. Most recently, personalized Web search systems [2] emerged as the most active area of research in user modeling and personalization.

What makes user models attractive in a Web search context is their ability to separate relevant and non-relevant documents according to user information needs. Observing user actions over time, a personalized search system attempts to capture user interests at the task level and to represent it in the user model. The personal difference or task context can then be applied to the information provided by the systems.

Unlike text-based personalized search systems that make use of ranked lists, spatial visualization approaches expand the one-dimensional information space to 2- or 3- dimensional representations of retrieved results, appealing to human intelligence and human ability to analyze visual information. We have tried to combine the ideas of user modeling and visualization to develop a "best of both worlds" approach for assisting users in their search process. We explored a promising technology that applies user modeling in the context of spatial relevance visualization, so that the personalized search task could be more interactive and exploratory. Our pilot studies demonstrated that the user model's ability to separate relevant and non-relevant documents could provide substantial assistance in an information visualization context. In this paper we briefly present our adaptive visualization technology and the evaluation result.

2. VISUALIZAING USER MODEL AND USER QUERY USING VIBE FRAMEWORK

In personalized search systems, three important elements compose the whole framework: query, user model, and the ranked list. The users themselves provide the queries, and the system constructs their user models by observing their actions. Among various approaches, content-based user models are composed of important keywords extracted during the user interactions and can represent user interests. The system can return personalized search results with these two input variables.

We tried to visualize these three elements using an application named VIBE (Visual Information Browsing Environment) that was originally developed by Olsen et al. [3]. We extended it to implement an adaptive information access on top of the original framework. VIBE is a reference point – called POI (Point of Interest) – based spatial visualization system. The entities to be visualized are allocated on a 2-dimensional space, and their locations are determined by the similarity ratios to the POIs. For example, if the similarity values of an entity between POI_a and POI_b are 0.8 and 0.4, then the entity is placed in a position two times closer to POI_a than POI_b because the similarity ratio is 2:1. Therefore, entities are placed closer to the more similar POIs on the screen. Users can drag the POIs interactively, and the entity positions are updated on the fly following related POIs.

We defined two types of POIs: query POIs and user model POIs. Query POIs are simply query terms that were entered by users. User model POIs are the top N most important terms from the user model. The locations of the documents are decided by their similarity ratios to the POIs. Because the documents can be represented as term vectors (bag of words approach) [4], the similarity values between the documents and the POIs (query terms or user model terms) can be easily calculated using standard similarity measures, such as cosine coefficient.

Figure 1 shows an example of this visualization. Two query terms (CONVICT and PARDON painted in yellow) and the five most important user model terms (YEAR, POPE, ESPIONAGE, PRISON, and RUSSIA painted in blue) are displayed as POIs. Squares represent retrieved documents and are aligned according to their similarities to the query terms and the user model terms. For example, the squares distributed closer to PARDON represent the documents that are more related to the query than the user model and are more related to PARDON among two query terms. Unlike the original VIBE that aligned the POIs on a circle by default, we added two new default layouts: Hemisphere and Parallel. Figure 1 shows the Hemisphere layout and Figure 2 shows the Parallel. This new idea was motivated by the fact that we have two different group of POIs – query and user model POIs – and need a means to visually separate them in the information space.

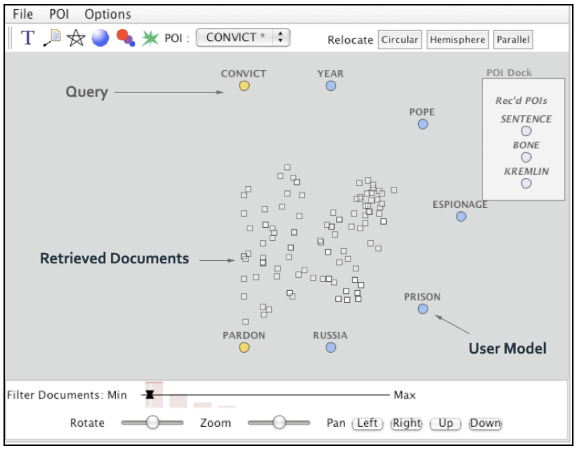

Figure 1: Adaptive VIBE Visualization (in Hemisphere layout)

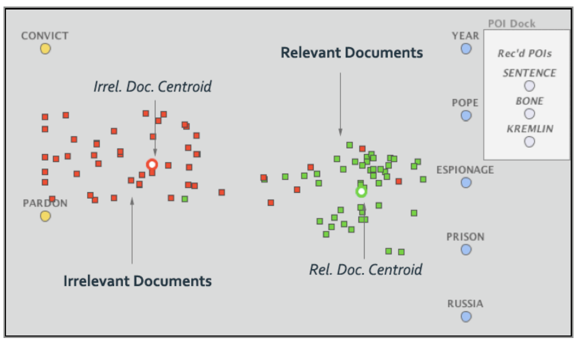

Figure 2: Adaptive VIBE Visualization (in Parallel layout) The simulation result shows a separation of relevant and non-relevant information painted in green and red squares respectively

3. SIMULATING ADAPTIVE VISUALIZATION WITH REAL DATA

In order to test the effectiveness of the proposed adaptive visualization approach, we extracted log data from our previous study, which is a personalized search study based on a simple textual interface, and simulated the visualization with the data using the adaptive VIBE framework. We extracted three important categories of data from the log: queries, user models, and the documents returned by the system.

We adopted a special type of user model for this study, which is called a task model. It is equivalent to the conventional user models in that it reflects user interests, but is more focused on specific tasks. The subjects from the previous study were asked to solve a specific task by entering queries, retrieve documents, and save important textual passages they found using the personalized search system. The system automatically extracts important terms from the user-selected passages and constructs a user model. Therefore, a user model consists of a set of terms. Each user model term is weighted according to its importance (frequency). Using the user query and the user model constructed in this way, the system provides personalized search results to the subjects.

We extracted 105 transaction records from the log data and fed them into the proposed visualization framework as shown in Figure 2. One important visual cue in this figure is the color of documents (squares). By matching the topics the subjects were asked to work on and the ground-truth data manually tagged by humans beforehand, we were also able to mark each document as "relevant" or "not relevant". Therefore, the document color represents this relevance information (green = relevant, red = non-relevant). The relevance information was never visible to the subjects during the previous study but was added later to the simulation for the authors to observe the effect of the user model.

4. THE USER MODEL EFFECT TO VISUALIZATION

As shown in the simulation example (Figure 2), we can easily observe that there is a separation between green and red squares, i.e., relevant and non-relevant documents to the corresponding topic. This effect of separation was observed from many of the simulation results among the 105 log records. In order to investigate this separation numerically, we calculated the centroids (marked as green or red circles in Figure 2) of the relevant and non-relevant document clusters and then compared their horizontal positions. The centroid positions were achieved by averaging the document positions within the corresponding clusters. The reason why we compared the horizontal positions, but not the vertical, was that the query and user model POIs were separated horizontally. Table 1 shows the result. The average horizontal positions of the relevant and non-relevant document clusters were presented in pixels on a Java Swing plane where the origin ($x = 0$, $y = 0$) is the left topmost position. The average positions of the relevant document cluster centroids were greater than the non-relevant document clusters for all three layouts, suggesting that the relevant documents are positioned closer to the user model. The differences were all statistically significant (paired Wilcox signed rank test) and the separation was greatest with the Parallel layout.

Table 1: Comparison of Cluster Centroid Positions

POI Layout	Circular	Parallel	Hemisphere
Relevant	304.3	300.9	337.7
Non-relevant	283.9	207.96	295.3
Difference (Relevant – Non-relevant)	20.4 ($p < 0.01$)	92.94 ($p < 0.01$)	42.40 ($p < 0.01$)

5. CONCLUSIONS

In this study, we proposed an adaptive visualization framework by using an extended version of VIBE. The simulation of the proposed framework using the log data extracted from a personalized search system suggests that our adaptive visualization framework can spatially separate relevant and non-relevant information. This separation will be able to help users analyze and filter huge volumes of information.

6. REFERENCES

[1] S. Gauch, M. Speretta, A. Chandramouli and A. Micarelli, User profiles for personalized information access. in: The Adaptive Web: Methods and Strategies of Web Personalization, P. Brusilovsky, A. Kobsa and W. Neidl,

eds., Lecture Notes in Computer Science 4321, Springer-Verlag, Berlin Heidelberg New York, 2007, pp.54-89.

[2] A. Micarelli and F. Sciarrone, Anatomy and empirical evaluation of an adaptive Web-based information filtering system. User Modeling and User Adapted Interaction, 14(159-200) (2004).

[3] K. A. Olsen, R. R. Korfhage, K. M. Sochats, M. B. Spring, and J. G. Williams, "Visualization of a document collection: the vibe system," Inf. Process. Manage., vol. 29, no. 1, pp. 69-81, 1993.

[4] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," Commun. ACM, vol. 18, no. 11, pp. 613-620, November 1975.