Visualization of a user model in educational document retrieval

Swantje Willms

Department of Information Science and Telecommunications,
University of Pittsburgh, Pittsburgh, PA, USA
swillms@mail.sis.pitt.edu

Abstract. We propose to visualize the user model in relation to the documents that are relevant to the user's task. Each component of the user profile is represented by a point of interest in a reference-point based visualization. We employ this visualization in a learning environment where its purpose is to help the student in identifying relevant documents for study.

1 Introduction

We propose to combine information retrieval (IR) and user modeling by visualizing the user model in relation to the documents that are relevant to the user's task. We address the workshop question: How can the user use the learned user model? The user is aided in using the model by giving her visual access to the relationship between her interests and knowledge (the components of the user model) and the retrieved information (documents).

In our research, we focus on the user model in a learning environment. Students have access to valuable resources outside of the lecture notes and required readings. A reason that these are often not accessed is that it is cumbersome for the student to assess their immediate relevance. When the user model is represented in a vector based representation, it can directly be visualized in a reference-point based visualization. In a spatial visualization like WebVIBE, each user profile component can be represented as a separate reference point (e.g. interest, knowledge) among other reference points (e.g. lecture, query).

The visualization can help the user to identify documents that they want to explore, in particular its purpose is to help the student identify relevant documents for study.

2 Related Work

In most information retrieval systems the query or information need is the only aspect of the user that is represented. No real user models are employed.

A need to employ user modeling techniques for visualization systems has been recognized [4]. A couple of experimental tools for visualizing user models have been proposed [5, 11]. They focus on visualizing the user model itself, whereas my work focuses on the relationship between the user model and the documents. Spatial visualization has been researched for IR tasks, but reference points have mostly been query terms and similar items. Even though the user profile is stated as an obvious example of a reference point by Korfhage [6, p.163], there hasn't been much research on this aspect.

We want to include the user model in the visualization: The goal is to help the students to identify documents that they want to study taking into account the lecture topic on one hand and what is known about the student, i.e., the user profile, on the other hand.

In the context of the learning environment we are exploring, a map based visualization called KnowledgeSea [1] has been used. KnowledgeSea visualizes all the concepts associated with a class and the underlying whole collection in a learning environment. It displays a two-dimensional map of the educational documents, with a set of keywords in each cell that describe the documents in this cell (Figure 1). Documents that are semantically related are close to each other on the map. If they are in the same cell, they are considered very similar.

3 Representation and Visualization

The vector model provides a consistent representation of all components as vectors, i.e., of the documents, the user model components, and any other points of interest.

3.1 Vector Representation

In the vector model of information retrieval [7], indexing terms are regarded as the coordinates of a multidimensional information space. Each document in the vector model is described by a vector of term values.

A reference point or point of interest (POI) is any definable point in the document space; it is a point or concept against which a document can be judged. A reference point is defined by a set of weighted terms, represented by a vector of term values.

The meaning of the value or weight differs slightly depending on the kind of item that is represented by the vector, i.e., document or POI (see section 4), but a higher weight always represents something like higher importance or relevance.

The user model in our research is represented as vectors of user profile components. Section 4.1 describes how the user profile components are acquired. The purpose of the visualization of user profile components as points of interest is to help the student to identify relevant documents for study. In the past, relevant documents for the student have been shown as a list, but a multi-dimensional display may be easier to understand. It is more flexible because it can show how the potential study documents relate to the POIs.

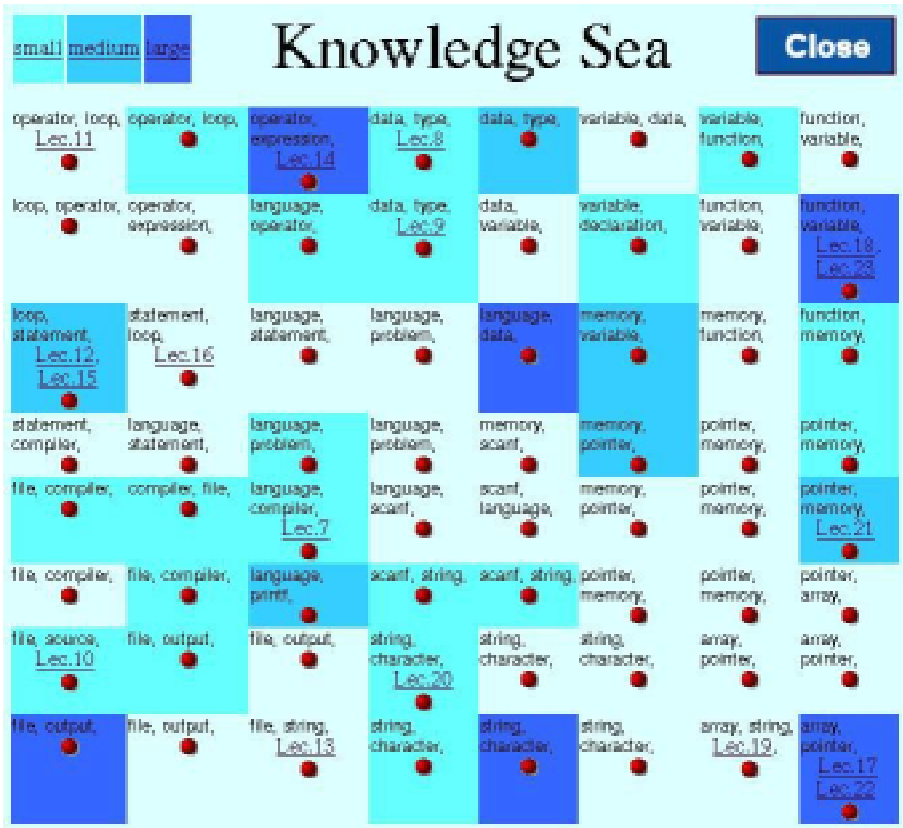

Fig. 1. KnowledgeSea map for “Introduction to Programming” class

3.2 Spatial Reference-Point Visualization

In a spatial visualization, each document is displayed as a point in the document space and placement of documents is based on attraction according to similarity to the reference points. Overall trends about POI-document distribution in a set of documents can emerge from a spatial display. Spatial visualizations explicitly visualize varying degrees of similarity.

The VIBE (Visual Information Browsing Environment) [8] spatial reference point visualization interface focuses on presenting the document as a single entity within a document collection. VIBE places the points of interest at the vertices of its display. Individual documents are represented as icons [9]. VIBE is based on the ratio of similarity measures for a document with respect to multiple reference points. The user places the icons that represent the reference points on the screen where a reference point can be any attribute to which a numeric association strength can be assigned. A weighted centroid model (spring model) is then used automatically to place the document icons according to the strength of their relationships to the reference points. Therefore, the position doesn't indicate the absolute significance of a document, but instead the relative significance of the reference points for the document. The reference points can be arbitrarily located and moved on the two dimensional space which helps disambiguate the locations of documents. Showing the documents with respect to multiple reference points allows the users to perceive relationships of each document with respect to all of those reference points.

In WebVIBE [12], the Java version of VIBE [8], the POIs are represented by magnets as a metaphor for attraction (Figure 2); as in the original VIBE, they define a coordinate system on the display to present a virtual document space. Each POI in WebVIBE is based on a vector of keywords representing a user profile component or other POI (see section 4). The structure of the presentation is user-defined in WebVIBE as well, because the users can change the display interactively by selecting and placing the POIs.

4 Components and Points of Interest

The POIs for the display will be generated behind the scenes based on the criteria described below and used for a WebVIBE display. Each POI consists of a term vector. The following sections describe how the different POIs are derived and represented.

4.1 User Model

The user model will consist of components including the student's interest and the knowledge of the student, each treated as a separate POI in the visualization.

The students express their **interest** by highlighting the part of a document that is relevant to them or by rating the document for relevance. In a process similar to that used for WebMate [3], the highlighted parts (or the positively

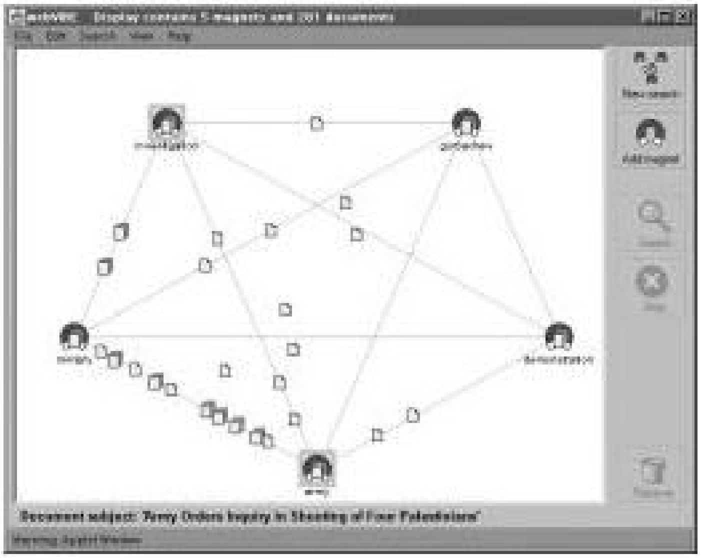

Fig. 2. A sample WebVIBE display

rated document) are preprocessed by deleting stop words and stemming, and possibly extracting titles to give them more weight. The frequency vector for the document is extracted. Then it is either added to a set of vectors that represent the student's interests or combined with the one most similar to it.

Student **knowledge** is represented by weighting keywords (from unknown to known) based on test scores. Students will take a quiz after a lecture topic has been covered in class (this evaluation could also be done on a voluntary basis). Each multiple choice question in a quiz is associated with some terms or keywords (extracted from a paragraph from lecture material) that correspond to the concept(s) addressed by the question. If the student answered the question correctly, it is assumed that she knows the concept represented by the keywords. Answering several questions with the same keyword correctly increases the weight for a keyword. This increased weight represents increased knowledge (up to a maximum).

Future work may include an explicit representation of student **needs** based on test scores. In addition, group models of the class interest or the interests of groups of students could be incorporated as POIs.

4.2 Other POIs

Apart from the user model, another important POI, the **lecture topic**, is represented by keywords extracted from the corresponding lecture notes document (a set of slides). A non-homogeneous lecture will be split into two or three topics if necessary.

In future work, other POIs such as those representing a traditional **query** may be included as an option for the display.

4.3 Similarity

The similarity between the POIs and the documents is computed using a similarity measure such as the cosine measure. The cosine measure has been widely used for its simplicity and effectiveness [10]. The position of each document within the WebVIBE display is based on the ratio of the similarity measures for a document with respect to the POIs. A weighted centroid model places the documents according to the strength of their relationship to the POIs.

5 Research Design Overview

Our system shows the relevant documents in a multidimensional similarity-based display that shows relationships to the user model. The visualization displays the relationship of the study documents to the student's **knowledge**, **interest**, and the **lecture topic**. We will explore two hypotheses. The first hypothesis is that the students may be able to find relevant documents faster when they have a visualization of their interest and knowledge available and use it to identify relevant study materials. The study will determine whether there is an improved success rate in finding relevant documents within the allotted time frame vs. a control condition using presentation in lists, i.e., without visualization and user model. A second hypothesis is that students are more likely to make use of external material when they can judge its relevance with the help of the visualization. We will conduct two kinds of experiments involving real users, the students enrolled in a class, to explore these hypotheses: 1. Long term observation and 2. Controlled sessions. In addition, the students' satisfaction with the visualization will be assessed through a questionnaire.

5.1 Resources

Some of the needed material already exists for classes such as "Introduction to Programming". The existing lecture notes in form of slides build the internal document collection. In addition, the students have access to a couple of online tutorials as external resources. These can be referenced by page, the pages for one tutorial making up one collection.

5.2 Visualization

Different "sections" of a class will have different conditions, e.g. one with WebVIBE visualization, one without (control condition). WebVIBE will be employed in the context of the KnowledgeSea visualization [1] described earlier in this paper. KnowledgeSea provides an overview of the keywords associated with the educational documents. On the two-dimensional map of the educational resources that builds the core of KnowledgeSea (Figure 1), each cell displays a set of keywords associated with the documents located in this cell. Usually, the user would be linked to a list of documents from the cell. Instead the users will be linked to a WebVIBE display (Figure 2) so that they can see how these documents relate to each other with respect to the POIs.

The WebVIBE display the student is presented with will have at least the three POIs: **interest**, **knowledge**, and **lecture topic** as described above. The student will want to learn about concepts that are far away from her knowledge and close to her interest or the topic. The centroid of the lecture can be seen as a target because it represents what the student is supposed to know. Therefore, the student may want to explore those documents that relate to both their interest and the lecture topic, and also look at documents that integrate these with her previous knowledge, but not at documents that are related to knowledge only.

5.3 Experiments

Over the course of the semester, we will keep track of the documents that the students access. Using these data, we can find out if students are more likely to access external material if they access the documents through the visualization system.

In addition, we will schedule a session of specified length in which the student is asked to identify study material relevant to a specific topic. In this controlled session, we can assess whether students find relevant documents faster with the visualization. At the beginning of this focused session, the student will state her goal in natural language; a list of keywords associated with this goal could be used as a starting point for the student's interest. In this manner, the student's perceived goal will be adjusted (like the model for the user's objective in METIOREW [2]).

6 Conclusions and Future Work

In this paper, we present a way of visualizing the user model in the context of educational document retrieval. The visualization represents each user profile component as a reference point. The information is delivered to the users so they can assess the relationship between the documents and the user model. Currently considered POIs are an individual user's interest, knowledge and the lecture topic.

Additional conditions that one could incorporate in a future study include the following: One could extend the user model by (1)(a) the **class interest** as a group model extracted from the interests of individual students or (b) several separate group models based on "categories" of students such as good, average, and poor students; (2) an explicit representation of student **needs** based on test scores weighted inversely to the representation for knowledge (i.e. low scores get high weights) and on topics that have been covered in class, but which the student has never accessed. This POI could prove especially useful for review sessions.

In addition, different ways of acquiring and updating this user model and its influence on the visualization can be explored as long as it can be represented as term vectors. The visualization could provide additional or alternative POIs such as the class/group interest and needs just mentioned or a POI based on a traditional keyword query.

One could add additional document collections such as a glossary.

The WebVIBE visualization could be employed in a stand-alone fashion without the context of the map based KnowledgeSea visualization.

Acknowledgments

I would like to thank Michael Lewis and Peter Brusilovsky at the Department of Information Science and Telecommunications at the University of Pittsburgh for their support with this project.

References

1. Brusilovsky, P., Rizzo, R.: Map-Based Horizontal Navigation in Educational Hypertext. Journal of Digital Information **3(1)** (2002)
2. Bueno, D., Conejo, R., David, A.A.: METIOREW: An objective oriented content based and collaborative recommending system. Third Workshop on Adaptive Hypertext and Hypermedia (AH2001) (2001) 310–314
3. Chen, L., Sycara, K.: WebMate: A personal agent for browsing and searching. 2nd International Conference on Autonomous Agents (Agents'98) (1998) 132–139
4. Grawemeyer, B.: User Adaptive Information Visualization. 5th Human Centred Technology Postgraduate Workshop, University of Sussex, School of Cognitive and Computing Sciences (HCT-2001) (2001)
5. Kliger, J.: Model Planes and Totem Poles: Methods for Visualizing User Models. Master of Science, MIT, Media Lab (1995)
6. Korfhage, R.R.: Information Storage and Retrieval. New York: Wiley (1997)
7. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Communications of the ACM **18** (1975) 613–620
8. Olsen, K.A., Williams, J.G., Sochats, K.M., Hirtle, S.C.: Ideation through visualization: the VIBE system. Multimedia Review **3(3)** (1992) 48–59.
9. Olsen, K.A., Korfhage, R.R., Sochats, K.M., Spring, M.B., Williams, J.G.: Visualization of a document collection: The VIBE system. Information Processing and Management **29** (1993) 69–81
10. Salton, G., McGill, M.J.: Introduction to Modern Information Retrieval. McGraw Hill (1983)
11. Uther, James, Kay, Judy: VIUM: A Web-Based Visualisation of Large User Models. Ninth International Conference on User Modeling (UM'03) (2003)
12. Homepage for WebVIBE. URL: http://www2.sis.pitt.edu/~webvibe/WebVibe/