# Supporting Document Triage via Annotation-based Multi-Application Visualizations

Soonil Bae[4], DoHyoung Kim[1], Konstantinos Meintanis[1], J. Michael Moore[1],
Anna Zacchi[1], Frank Shipman[1], Haowei Hsieh[2], Catherine C. Marshall[3]

[1]Dept. of Computer Science
Texas A&M University
College Station, TX 77843

[2]School of Library &
Information Science
University of Iowa
Iowa City, IA 52242-1420

[3]Microsoft Research
Silicon Valley
1065 La Avenida
Mountain View, CA 94043

[4]Samsung Techwin
13th Fl., KIPS Center
Yeoksam-dong Kangnam-gu
Seoul 135-980, Korea

soonil.bae@samsung.com, shipman@cs.tamu.edu, haowei-hsieh@uiowa.edu, cathymar@microsoft.com

## ABSTRACT

For open-ended information tasks, users must sift through many potentially relevant documents, a practice we refer to as document triage. Normally, people perform triage using multiple applications in concert: a search engine interface presents lists of potentially relevant documents; a document reader displays their contents; and a third tool—a text editor or personal information management application—is used to record notes and assessments. To support document triage, we have developed an extensible multi-application architecture that initially includes an information workspace and a document reader. An Interest Profile Manager infers users' interests from their interactions with the triage applications, coupled with the characteristics of the documents they are interacting with. The resulting interest profile is used to generate visualizations that direct users' attention to documents or parts of documents that match their inferred interests. The novelty of our approach lies in the aggregation of activity records across applications to generate fine-grained models of user interest.

## Categories and Subject Descriptors

H.3.3 [**Information storage and retrieval**]: Information search and retrieval – *Search process*

## General Terms

Design, Experimentation, Human Factors.

## Keywords

Document triage, multi-application user modeling, visualization.

## 1. INTRODUCTION

The Web is a vast information resource that may be brought to bear on many different types of activities. Some activities dictate a specific information requirement, such as the location of a restaurant, the value of a stock, or the answer to a question. We are concerned with more open-ended information gathering tasks—analysis tasks in particular—in which people collect Web documents for interpretation and synthesis.

Even with the best search engine and the most effective query formulation, such tasks require people to work through long lists of documents to synthesize the information they need; there is usually no single document containing one right answer. In fact, as people skim early documents, they may determine additional information needs that suggest further queries and result in even more documents to process [3]. We call this rapid assessment of documents based on their potential value information triage [19] or document triage [2]. Document triage involves selecting which documents to examine in more detail; identifying the most useful parts of documents; and keeping track of progress through the search results.

A system can support document triage by recommending the documents that best match a user's interests, thereby ensuring that the user's time is spent efficiently on the most relevant documents. Many techniques and sources of evidence may be used to generate these recommendations. Recommendations may be based on the interests, activities, and outcomes of similar information tasks as they are performed by other users; they may be based on metadata similarity or on a network analysis of relationships among documents. In the work we present in this paper, recommendations are based on demonstrated user interest; in other words, the user's previous interactions with the document collection, along with the characteristics of the documents, are used to infer the user's interests.

For over a decade, we have developed spatial hypertext workspaces that support information analysis and document triage [1][2]. These workspaces accelerate document assessment by providing users with lightweight methods for expressing the perceived value of documents and the relationships among them.

But assessment often entails skimming and selective reading, which usually occurs in a separate reader, such as a Web browser or a specialized reading tool such as Adobe Acrobat or Microsoft Word. Thus, users often interact with multiple applications during triage: for example, they may consult a Web browser that displays a list of search results; they may use a reader to examine the content of individual documents; and they may use a note-taking tool to keep track of what they have read. To better support triage, we have adopted a strategy of inferring user interest based on user interaction with multiple triage-related applications [1]. Our past research shows that combining evidence from an information workspace and a document reader improves the interest model.

This paper describes our efforts to use user interest models as a basis for visualizations that draw a user's attention to similar documents and document parts. In particular, we discuss how user

*JCDL'10,* June 21–25, 2010, Gold Coast, Queensland, Australia.
Copyright 2010 ACM 978-1-4503-0085-8/10/06...$10.00.

interactions with documents — including implicit expressions of interest (e.g. scrolling), along with explicit expressions of interest (e.g. annotations) — can be combined with document and content characteristics and translated into visualizations which may be applied to other relevant documents to simplify the triage process.

The paper describes an extensible architecture that supports document triage and the three software tools we modified or developed to address different aspects of the triage task: a spatial hypertext workspace; an annotation plug-in for Firefox; and an interest profile manager. We begin by surveying related work and presenting our approach and architecture. We then describe the three tools in detail. Finally, we discuss how the tools work in concert to generate the triage visualizations and present the results of a study evaluating the effectiveness of our approach.

# 2. RELATED WORK

Related work falls into three main categories: (1) investigations of document triage in the field and in the lab; (2) research into methods of identifying and modeling user interest; and (3) exploration of visualization techniques to focus user attention and to aid in navigation through complex information spaces.

## 2.1 Document Triage

Document triage has been studied from at least three different perspectives in the field and in the lab. The first perspective centers on the documents, modeling how human interest varies with document characteristics: titles, length, embedded images, and other features that may be used to distinguish documents [8]. The second perspective centers on what people do with documents, how they interpret, structure, and categorize them in a task context [2][19][22]. A third perspective investigates the effect of information technologies and digital media on the process: how does triage on paper differ from triage using electronic media and tools [4]. For example, does the ability to construct complex category hierarchies make people focus on organizing rather than reading? An early triage study comparing the use of spatial hypertext with paper found that it did [19]. Moreover, a subsequent study found that display configuration had a significant effect on triage; using multiple displays resulted in increased transitions between activities [2]. In other words, document triage is sensitive to the affordances of the medium.

## 2.2 Interest Modeling

A second category of related work arises from the need to identify documents of interest. Interest modeling may either be based on explicit indications of user interest (e.g. ratings), implicit interest indicators (e.g. click-through records), or a combination of the two. We examine related work in each area.

### *2.2.1 Explicit Indicators*

Explicit interest indicators rely on users identifying valuable documents, e.g., by assigning ratings. Several digital library systems use this approach [20][23]. Because they are user-assigned, explicit indicators are easily understood and require no further interpretation. However, eliciting this information can interfere with a user's normal reading and browsing patterns [7]. Users may also stop rating documents when they do not realize an immediate benefit from their efforts [14]. Moreover, users rate far fewer documents than they read [24]. Thus the benefits of explicit indicators may be offset by their drawbacks.

### *2.2.2 Implicit Indicators*

Implicit interest indicators are based on users' actions rather than on explicit value assessments. During triage, readers indicate their interest in documents by how they interact with them: by how much of the document they examine (e.g. how far into a document they scroll); by annotating content (e.g., highlighting passages); by how they categorize the document (e.g., stacking it with other interesting documents); and through other behaviors that in part rely on the tools they are using. This interest may be recorded as users interact with documents via the triage software; these documents may be characterized via feature extraction.

But it is insufficient to simply characterize documents and record users' interactions with them; it is also necessary for triage tools to interpret the interactions' meaning to arrive at a profile of the user's interests. The meanings of interactions may vary among applications; they not only offer different interactive functionality, but also other factors, such as ease of use, confound which software features are used.

Systems may model implicit interest using knowledge structures, which can be pre-defined [5][11] or created on the fly, based on user behaviors and document features. Some models incorporate characteristics of the larger user population [6][17]. Models that focus on individual users can be task specific [21] or can adapt as activity is observed over multiple sessions [13][15]. User behaviors may be difficult to decipher and use since they can be interpreted in different ways. User expression (e.g. annotation) may provide more focused input for user models with fewer opportunities for misinterpretation [12][29].

These systems have several limitations if we look more closely at the characteristics of documents and of the triage task. First, these systems generally treat documents as an atomic unit. However, useful documents may be long, and cover multiple subtopics; users may read some segments and ignore others. Interest profiling systems may be enhanced by recording which document portion(s) pique the user's interest. Second, these systems monitor user activity within a single application. Real triage may take place using multiple applications (e.g., Firefox and Acrobat). Recognizing this, we found that models combining interest information from multiple applications are more effective than those that rely on information from a single application [1].

## 2.3 Visualizations Aiding Navigation

For users to take advantage of system-identified "interesting documents," they must be made aware of these documents. Several strategies have been used successfully to focus the user's attention on these documents: (1) the presentation order may be changed to reflect the interest profile; (2) the display may be animated to focus the user's attention; (3) the visual characteristics of the document surrogates or passages within the document may be changed to call attention to them.

As an example of the first strategy, user models may be used to re-rank the documents returned by a standard query [21]. As an example of the second strategy, a hierarchical display of document surrogates can be zoomed in and out to focus the user's attention; document surrogates can also be hidden to reflect inferred user interest [11]. The third strategy does not rely on changing the order in which documents are presented or changing the user's vantage point on them; rather, it uses visual characteristics to highlight the documents or parts of the documents in place. For example, implicit queries in Data Mountain identify documents similar to the one being viewed; these query results are brought to the user's attention by outlining the documents of interest in green [10]. XLibris draws a user's attention to passages of potential interest (based on query rank) by using color and icons to highlight them in a document overview [29]. In this work, we use visualizations that parallel, but are visually distinct from, the kinds that users create in the tools.

## 3. APPROACH & ARCHITECTURE

This research examines how inferred models of user interest may be used to generate visualizations to aid the user in performing document triage. In particular, we explore how applications that are normally used for reading and organizing documents can contribute to and share these models to mutually support the activity with different forms of visual feedback.

We use a four step approach to generate visualizations:

- Users' interests are inferred from their interactions with a collection of relevant documents (for example, if the user assigns the same color—say, red—to nodes representing two documents A and B, these interactions are recorded and are taken as indications of the user's interest in A and B);
- Classes of user interest are identified and represented based on relationships among the documents of interest (in this step, the similarity between the content of documents A and B is characterized as a class of user interest);
- Documents similar to one or more classes of interest are selected (if document C has content that is similar to documents A and B, it is assumed to be in the same interest class as documents A and B); and
- Visualizations are generated to reflect how these documents are related to the inferred interests (the visualization the user applied to documents A and B partially determines the visualization applied to document C; in this case, the node representing document C would be highlighted in red).

People often use three types of applications to perform triage on a collection of Web-based documents: an overview application (to specify queries and examine the results); a reading application (to examine and possibly annotate the individual documents); and an information management application (to organize, manage, and interpret the documents). In the case of Web documents, these three types of applications might correspond to a browser that displays search engine results, a PDF/html viewer that displays individual documents, and a text editor that enables the user to gather, organize, and manipulate valuable URLs.

Our architecture must handle these types of applications to support basic triage activities. It must also include an Interest Profile Manager so that it can accumulate and analyze implicit and explicit interest indicators from the constituent applications and generate the appropriate visualizations for each. Finally the architecture must have a provision for storing interest profiles so they persist across sessions. Applications must each engage in two-way communication with the Interest Profile Manager in an extensible way so that additional applications can be added as appropriate; e.g., new reading applications may be added to the architecture to extend the capabilities offered by our proof-of-concept Web page reader. Such extensions will enable the user to work with new content types.

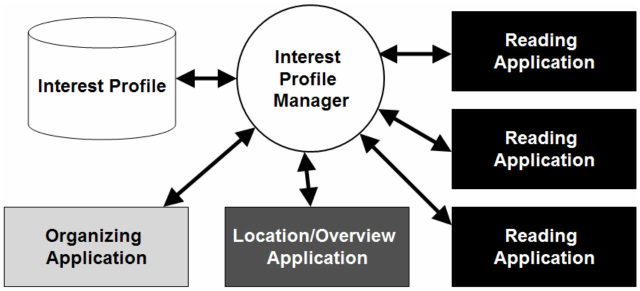


**Figure 1. Communication between triage applications**

Figure 1 shows this overall architecture. In the implementation we discuss in this paper, we simplify the general architecture by collapsing the application used to present the search results and the application used to organize them into a single application, the Visual Knowledge Builder (VKB). VKB has a built-in connection to Web search services and is designed to support organizing documents. The reading application is represented by WebAnnotate, which is intended as a proof-of-concept, rather than as a fully-functional reader. We have implemented the shared Interest Profile Manager (IPM) and a store for interest profiles; records of user activity in VKB and WebAnnotate are stored in the IPM and drive the visualizations that the IPM generates for either application.

The next sections describe the testbed applications, VKB 3 and WebAnnotate, focusing on the features that are important for triage. We then discuss the IPM and how it collects and analyzes user activity records to generate cross-application visualizations.

## 4. VKB 3: SYSTEM ENHANCEMENTS

The Visual Knowledge Builder (VKB) is a spatial hypertext workspace for collecting, analyzing, and organizing documents [27]. Much like its predecessor VIKI [18], documents in VKB are represented by *visual objects* that display metadata; these document surrogates are arranged in a hierarchy of two-dimensional spaces called *collections*. This paper introduces VKB 3, a major revision that extends the application's capabilities for working with Web information. We focus on a particular enhancement, a surrogate that represents Web documents using a combination of a visual thumbnail of the page and a display of its metadata; this new object type is called the *Web document object*.

VKB 3 also introduces the concept of *object layers*. Each VKB object can include multiple display layers. As shown in Figure 2, Web document objects have a main layer that can be configured to show various combinations of metadata (title, URL, and other properties) along with a thumbnail of the Web document, and a system layer. This layering enables a user to informally express interpretations of the documents (using color and other visual attributes) in the main layer, while the system

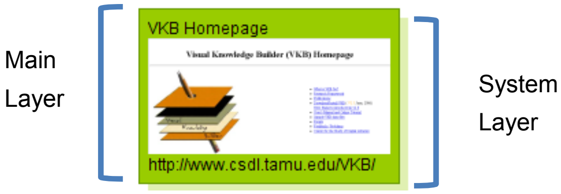

Main Layer
VKB Homepage
Visual Knowledge Builder (VKB) Homepage
http://www.csdl.tamu.edu/VKB/
System Layer

**Figure 2. Web document object with two object layers**

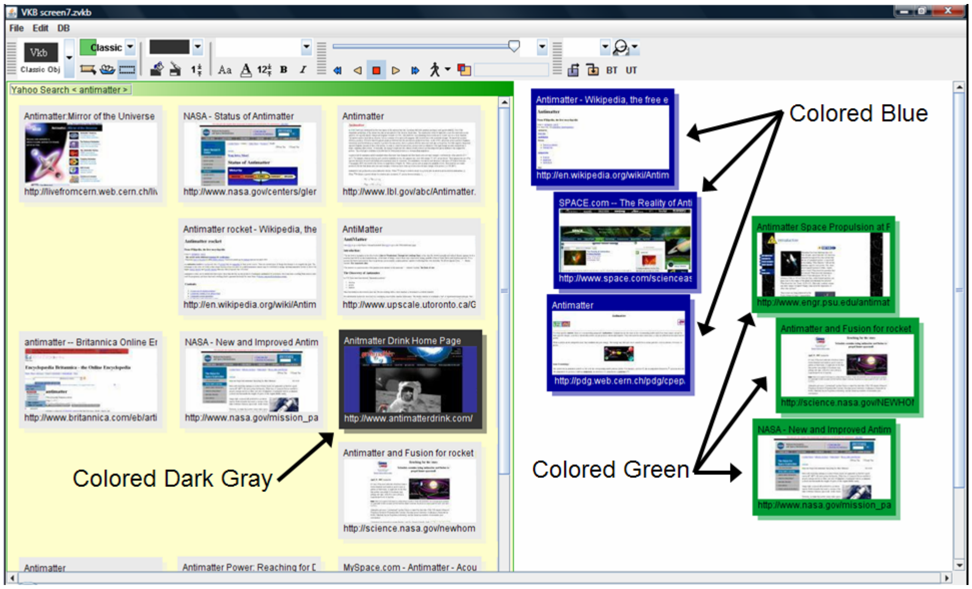

VKB screen7.zvkb
File Edit DB
Yahoo Search < antimatter >
Antimatter Mirror of the Universe
NASA - Status of Antimatter
Antimatter
Antimatter rocket - Wikipedia, the
AntiMatter
antimatter -- Britannica Online En
NASA - New and Improved Antim
Antimatter Drink Home Page
Antimatter and Fusion for rocket
Colored Dark Gray
Antimatter - Wikipedia, the free e
SPACE.com -- The Reality of Anti
Antimatter
Antimatter Space Propulsion at F
Antimatter and Fusion for rocket
NASA - New and Improved Antim
Colored Blue
Colored Green

**Figure 3. Web documents in VKB, presented as combinations thumbnails and metadata**

uses the system layer's visual attributes to provide cues, hints, or suggestions without interfering with user expression. We developed this extension in response to previous study results that suggested users were reluctant to express their own interpretations once the system had applied its own visualizations to objects [28].

Figure 3 illustrates a triage scenario. In our scenario, the user is speculating on the plausibility and utility of antimatter, starting from 20 Yahoo search results. The user has moved several articles he perceives as valuable definitions of antimatter and arguments for its existence into a pile and has colored them blue. He has created a green pile to categorize articles about antimatter's utility for space propulsion. An irrelevant Web page promoting an antimatter energy drink remains in the original collection ("Yahoo Search < antimatter >"); our user has changed the main layer for the Antimatter Drink Home Page to dark gray so he can easily ignore it. It is likely the workspace will grow to contain more collections as our user works with the documents and discovers the need for further searches; he may also create new collections to categorize documents and manage the space.

Prior versions of VKB are described in detail in [26][27][28]. The application is written in JAVA and runs on MS Windows, Apple's OS X, and Linux. VKB 2 may be downloaded from http://www.csdl.tamu.edu/VKB. VKB 3, shown for the first time here, will be released publicly after the beta release period.

For our triage investigation, VKB 3 has been extended to communicate with the Interest Profile Manager (IPM). VKB 3 sends two kinds of information to the IPM: records of user actions and the attribute/value pairs that characterize Web document objects in the workspace. In other words, as users open, move, color, delete, or otherwise modify document objects, records logging these actions are sent to the IPM; likewise, as document objects are added to the workspace or the attributes' values are edited, this information is also sent to the IPM. Table 1 lists the components of both forms of evidence of user interest. To ensure consistent reference to Web document objects, they are identified by their URL (Table 1 refers to this identifier as *Information ID*).

Communication with the IPM is two-way: not only does VKB send information to the IPM as input to the algorithm that computes user interests, VKB also receives information about user interests from the IPM, which it uses to modify the system layer of Web document objects in the workspace. Table 1 shows details about the information sent to and received from the IPM.

The IPM infers user interest based on information collected across all of the triage applications and sends the inferred interest back to VKB (as well as to the other applications). The IPM results are in three parts: information ID (so VKB can determine which object the interest applies to), interest classification (to specify topic), and interest level (to specify intensity). The IPM's algorithm for computing user interests and their intensity is discussed at greater length in Section 6, but it is important to stress that our technique for estimating interest is proof-of-concept and not one of the principal contributions of the research. After VKB receives the results of the IPM's analysis of user interests and their relative strength, the application translates these parameters into the visualization described in Section 7.

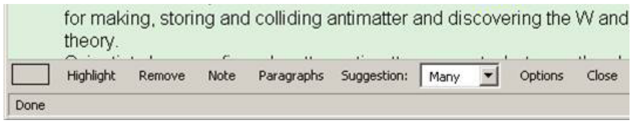
**Table 1. Information sent to and received from the IPM**

| **Information Sent to IPM** | | **Information Received from IPM** |
|---|---|---|
| User Actions | Attributes of Document Objects | |
| Create doc. object | Location | Information ID |
| Delete doc. object | Size | Interest classification |
| Move doc. object | Background color | |
| Resize doc. object | Border color | Interest level |
| Change background color | Border width | |
| Change border color | Font characteristics | |
| Change border width | Document title | |
| Change font/font color | Thumbnail info. | |
| Change document title | Document URL | |
| Change comment | Comment text | |
| Gain or lose focus | | |

## 5. WEBANNOTATE

When a user opens a document in the VKB workspace, it is displayed in the Firefox browser (this approach would work equally well with Microsoft's Internet Explorer). To further facilitate triage, we have developed a Mozilla Firefox add-on called WebAnnotate that provides basic annotation capabilities, collects data on users' interactions with documents, and uses interest data returned from the IPM to create visualizations designed to focus readers' attention on the portions of documents relevant to their interests. These visualizations will enable users to quickly locate what they want to read without taking the selected material out of context.

WebAnnotate supports several representative forms of annotation on HTML documents; once users activate the annotation toolbar (Figure 4), they can highlight text in different colors and can create colored sticky notes (editable translucent text boxes that can be moved anywhere on the HTML document). While sufficiently functional for our purposes, WebAnnotate's annotation model is not unique. In particular, Microsoft Word, Scrapbook, Annozilla, and Annotea were valuable references for designing WebAnnotate.



for making, storing and colliding antimatter and discovering the W and theory.

**Figure 4. Annotation toolbar of WebAnnotate**

WebAnnotate stores the reader's annotations separately from the HTML document, in the IPM, where they are used as input to the interest estimation algorithm. Whenever a user opens a HTML document, WebAnnotate checks the IPM for annotations to that document. If any are found, WebAnnotate regenerates them.

The communication between WebAnnotate and the IPM is similar to the communication between VKB and the IPM. Annotation information that is sent to the IPM includes the color and type of the annotation (whether it is a highlight or sticky note) as well as other terms necessary to reconstruct and describe the annotation (e.g. the anchor text or text of the note and the annotation's location). Our annotation representation assumes that documents are static, an assumption that reflects the nature of the triage task. The documents' DOM structure is used to specify the highlight's anchor location; likewise, sticky notes are reattached to Web pages according to their absolute (x, y) positions.

A prior triage study [1] showed that activity data, used in conjunction with document attributes, can be a meaningful source of evidence for inferring user interest. Thus WebAnnotate's second role is to collect detailed interaction data (e.g., scrolling, mouse clicks, and changes in focus), along with the characteristics of the corresponding web pages (e.g., length, number of embedded links, and number of images). This information (as specified in Table 2) is aggregated and sent to the IPM.

**Table 2. Document attributes and user events sent to IPM**

| Document Attributes | User Events |
|---|---|
| Document URL | Document URL |
| Creation time | Creation time |
| # of characters (length) | Click |
| # of links | Scroll |
| # of images | Focus in/out |
| Paragraph information | Time spent |

When a user opens a Web page, WebAnnotate extracts document attributes (see Table 2) and sends them to the IPM. WebAnnotate parses the text content into paragraphs and assigns paragraph IDs to them. This information is used by the IPM to infer and communicate potential interest on specific paragraphs to WebAnnotate. Unlike VKB, which sends events as they occur, WebAnnotate aggregates events until the user's attention turns elsewhere and the browser window loses focus. This local consolidation greatly improves communication efficiency.

The IPM communicates inferred user interests to WebAnnotate in a form similar to those sent to VKB; they are represented by information ID, interest classification, and interest level. Unlike VKB, the information ID consists of a document URL and a set of paragraph ids, because user interests are calculated at the paragraph level rather than the whole document level. WebAnnotate brings paragraphs to a user's attention by underlining them (i.e., users highlight; the system underlines). The IPM's user interest algorithm is described in Section 6. Section 7 discusses how the interest parameters are visualized.

# 6. INTEREST PROFILE MANAGER

The Interest Profile Manager (IPM) plays the central role in inferring user interest during document triage. As we have already discussed, the IPM collects information about interest-related activity from the triage applications. This information is aggregated and saved in the user's interest profile. Based on the interest profile, the IPM estimates user interest and broadcasts it to the participating applications.

Thus, the IPM acts as an interest profile server, while the triage applications act as interest profile clients. Any application that can be modified to include the interest profile client software interface can communicate with the IPM. Currently, VKB 3 and WebAnnotate include this interface. While both of these applications support two-way communication, this is not required – an application could merely provide information to the IPM or only receive interest information from the IPM.

## 6.1 Representing User Interest in Documents

Each application provides users with unique ways of interacting with information that can be used as a basis for inferring their interests. It is this type of interaction data that is sent to the IPM. For efficiency, applications may aggregate low-level activity data (such as scrolling data) before they send it to the IPM; major events then initiate communication with the IPM. For example, WebAnnotate sends aggregated data to the IPM when the document is replaced in the Web browser.

Although each application has unique information that may be used to gauge human interest, this interest assessment needs to be sharable among the different applications to be useful to the triage process as a whole. The IPM depends on an abstract XML representation for receiving interest-related information from applications and for broadcasting inferred interest to client applications. Because we realize that we cannot foresee all of the ways different applications will allow users to interact with documents, the representation is extremely general and extensible. Thus an interest profile consists of a document identifier, an application identifier, and a list of application-specific attribute/value pairs. In this way, new applications only have to inform the IPM of the attributes and how they demonstrate user interest when registering.

For example, an application like VKB would inform the IPM that the "move object" attribute, which counts the number of times a document object is moved, has a small positive impact on user interest while the "delete object" attribute has a strong negative impact on user interest. These effects are based on prior use of the tools [1]. Based on the registration, the IPM sets up the appropriate data structures to broadcast and receive interest information based on application's unique set of attributes and actions. In the current implementation, an Interest Profile is a list of documents and document segments, user activity associated with each document or segment, a term vector that characterizes each document or segment, and a set of visual and metadata features for each document or segment. User interests are computed as needed based on this data.

## 6.2 Inferring Classes of User Interest

The IPM uses the document attributes (e.g. metadata, term vectors, user-assigned color) to determine classes of user interest from the evidence of interest in individual documents. To aid in the creation of descriptions of document classes, the IPM includes term vector and metadata analysis capabilities as well as text tiling [16] capabilities to allow clients and the IPM to analyze text at the sub-document level. Currently, user-assigned color is used to identify the known members of an interest class and the other document attributes are used to characterize the document class that is used to select documents or document components that are similar to that class.

For example, consider the three dark blue VKB objects shown in Figure 3. The IPM will retrieve the documents (using their URLs), aggregate their term vectors, and identify the metadata that is common to all three. This processing results in a characterization of the document class. Next, the IPM will calculate an interest rating for each document. It does this by combining the user activity associated with each document across both triage applications. Finally, the three individual interest values are aggregated—currently they are summed—to quantify the user's interest in the document class.

## 6.3 Recognizing Similar Documents

Once the IPM has generated a list of document classes and calculated the user's relative interest in these classes, this information is broadcast to the registered applications and can be used to generate visualizations that distinguish the documents or parts of documents that fit into these categories. Because the current classes are based on how the user assigned colors (or other visual features) to objects in VKB and on how the user annotated with different colors in WebAnnotate, these classes have relatively intuitive mappings to the adaptive visualizations that are described in the next section.

# 7. VISUALIZATIONS

As we discussed in Section 6, the IPM provides a list of document classes to VKB and WebAnnotate. Each document class has a user interest estimation, an aggregate representation of the documents in the class, and feature(s) that distinguish the class.

This information is used to generate both within-application and across-application visualizations that may be applied to new objects or document segments. This section describes the four forms of adaptive visualization and illustrates them with examples derived from the antimatter scenario shown in Figure 3.

## 7.1 VKB to VKB

In the scenario, the user has classified some of the retrieval results in the workspace as either green, blue, or dark grey to reflect how (and whether) he will use them. VKB 3 uses the results of the IPM's analysis of these three implicit categories to propagate the likely color/classification to the remaining documents.

Currently, the system does this by comparing the term vectors for each document that the user has not interpreted yet with the established classes. If a document is close to one of the classes, the system layer's color changes so that it matches the document objects in that class. The color's saturation reflects the system's confidence (the degree of match between the document's term vector and the class term vector). Figure 5 shows examples of the resulting visualization: objects with green, blue, or gray system layers indicate that they are possible members of the three document classes; their shade indicates the strength of similarity. Objects for which the classification is ambiguous are unchanged.

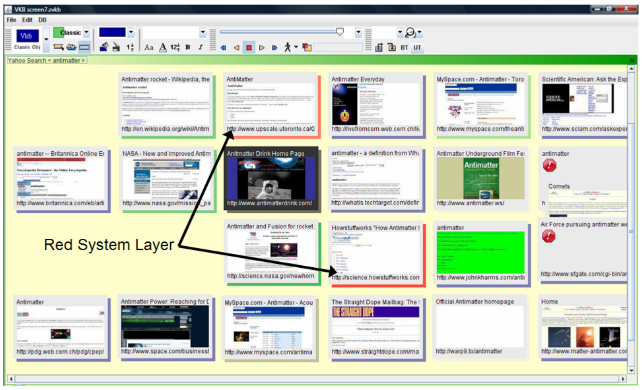
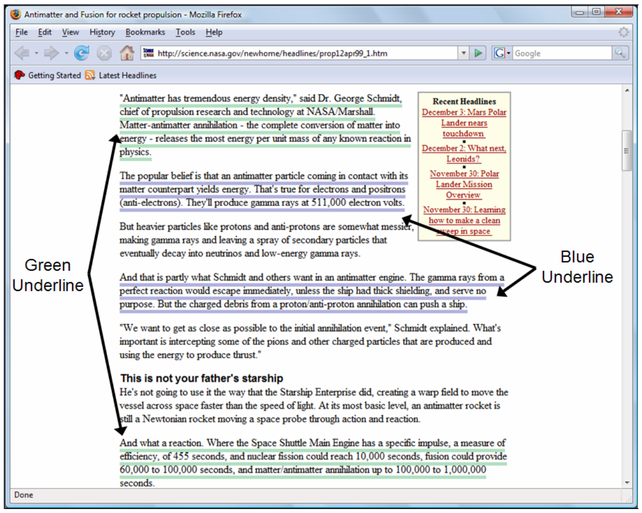


**Figure 5. Colors in objects' system layer indicate similarity to documents given that color previously by the user in VKB or similarity to document contents annotated in WebAnnotate.**

## 7.2 VKB to WebAnnotate

In VKB, color is applied to objects' system layers to reflect the whole documents' classification; WebAnnotate uses the same visual classifications to direct the user's attention to paragraphs of the documents that contain text that is similar to the class.


Antimatter and Fusion for rocket propulsion - Mozilla Firefox
File
Edit
View
History
Bookmarks
Tools
Help
http://science.nasa.gov/newhome/headlines/prop12apr99_1.htm
Google
Getting Started
Latest Headlines
Green Underline
Blue Underline
Done

"Antimatter has tremendous energy density," said Dr. George Schmidt, chief of propulsion research and technology at NASA/Marshall. Matter-antimatter annihilation - the complete conversion of matter into energy - releases the most energy per unit mass of any known reaction in physics.

The popular belief is that an antimatter particle coming in contact with its matter counterpart yields energy. That's true for electrons and positrons (anti-electrons). They'll produce gamma rays at 511,000 electron volts.

But heavier particles like protons and anti-protons are somewhat messier, making gamma rays and leaving a spray of secondary particles that eventually decay into neutrinos and low-energy gamma rays.

And that is partly what Schmidt and others want in an antimatter engine. The gamma rays from a perfect reaction would escape immediately, unless the ship had thick shielding, and serve no purpose. But the charged debris from a proton/anti-proton annihilation can push a ship.

"We want to get as close as possible to the initial annihilation event," Schmidt explained. What's important is intercepting some of the pions and other charged particles that are produced and using the energy to produce thrust."

**This is not your father's starship**

He's not going to use it the way that the Starship Enterprise did, creating a warp field to move the vessel across space faster than the speed of light. At its most basic level, an antimatter rocket is still a Newtonian rocket moving a space probe through action and reaction.

And what a reaction. Where the Space Shuttle Main Engine has a specific impulse, a measure of efficiency, of 455 seconds, and nuclear fission could reach 10,000 seconds, fusion could provide 60,000 to 100,000 seconds, and matter/antimatter annihilation up to 100,000 to 1,000,000 seconds.

Recent Headlines
December 3: Mars Polar Lander nears touchdown
December 2: What next, Leonids?
November 30: Polar Lander Mission Overview
November 30: Learning how to make a clean sweep in space

**Figure 6. The lightly-colored thick underlines indicate similarity of document segments to user interests.**

Applications can either perform their own similarity analysis between the classes of interest from IPM or ask the IPM to perform standard text segmentation and term-vector comparison. Because WebAnnotate is a lightweight add-on to Firefox, it has the IPM segment and compare the current document to the known user interest classes. The IPM tells WebAnnotate which paragraphs match, and the resulting text segments are given a thick light-colored underline to indicate the similarity. Figure 6 shows two such segments, one with a green underline, and two with a blue underline. These underlines may direct users' attention to relevant portions of the document they are reading; even if they were planning to read the whole document, the added underlines may cue them to the importance of the passage.

## 7.3 WebAnnotate to VKB

As users annotate documents in WebAnnotate, the annotation characteristics are sent to the IPM and aggregated with the user's annotations to other documents to create interest classes. Then, for each class, a term vector that characterizes the annotated text is broadcast to the triage applications.

As with VKB to VKB visualization, these term vectors are compared against those for the documents in the VKB workspace, resulting in the assignment of a similarity rating (not related, low, medium, or high) to each class. The border color of the document object's system layer is then assigned the color of the best matching class. The color saturation reflects the interest level; in other words, if the match is very close, the color will be opaque, and if the match is not as close, the color will be transparent.

Consider the case where our user continues his triage task by annotating documents on antimatter propulsion. He highlights passages relating to the design of antimatter engines in forest green and passages about antimatter production and its cost in red. This causes the IPM to add two new interest classes to his interest profile; one is represented by red, and the other by forest green. These classes are added to the three interest classes (blue, lime green, and gray) defined during VKB use.  Figure 5 shows the VKB visualizations that result from the reader's WebAnnotate annotations. The system layers of two objects are now red or light red because they are similar to the class containing documents about antimatter production and its cost.

## 7.4 WebAnnotate to WebAnnotate

If annotations reflect a user's interest in a particular concept, it may be valuable to see other discussions of the same concept later in the document or in other documents. XLibris introduced a similiar ability to propagate a reader's annotation ink throughout a document [25]. We generalize this technique so that WebAnnotate annotations are used to identify portions of newer documents similar to passages previously read and annotated.

As in the case of the visualizations that are propagated from WebAnnotate to VKB, the same interest classes are defined based on annotations' color, type and content. To identify segments of new or unread documents to bring to the user's attention, these classes are then compared against the segments of the document currently displayed in WebAnnotate generated by the text-tiling algorithm. When a match is identified, a thick underline of the appropriate color for the class is used to signal the similarity.

# 8. USER STUDY

We have performed a user study to evaluate the effectiveness of the VKB and WebAnnotate visualizations. The evaluation focuses on whether the recommendations help participants find documents of interest, whether the visualizations help participants keep track of their progress, and whether the new concept of layers helps reduce the conflict between user-authored visualizations and system-generated visualizations we identified in an earlier study.

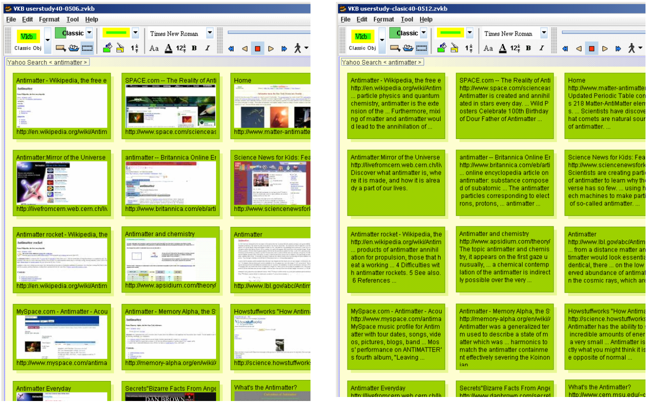
## 8.1 Experimental Design

The study was conducted in the Center for the Study of Digital Libraries at Texas A&M University. Twenty participants (14 graduate students and six undergrads, ranging in age from 18 to 39) were recruited via email. All use computers regularly and are familiar with searching and browsing the Internet. None had prior experience with VKB or WebAnnotate.

Participants were asked to select and organize documents on behalf of a science teacher who was preparing to teach a class on antimatter. We chose an unfamiliar topic so no participant would be a domain expert. Each participant started with 40 Web documents, the results of a Yahoo query on antimatter, which had been automatically placed in a grid in VKB. The instructions suggested that the task would take about 45 minutes, but that they could continue working as long as they needed to.

Participants were randomly divided into two groups with different software configurations. Participants in Group 1 were given the triage platform discussed in this paper: VKB 3 (with the enhanced visualization capabilities), the WebAnnotate plug-in to Firefox, and the Interest Profile Manager. Participants in Group 2 were given a simpler triage platform: VKB without the enhanced visualization capabilities (the VKB 2 release) and Firefox without the WebAnnotate plug-in; Group 2's triage platform did not include the IPM.

Figure 7 shows the initial document lists given to all participants; the workspace contained the same 40 Web documents in the same x,y arrangement. Group 1's document surrogates were thumbnails; Group 2's document surrogates displayed text snippets. Group 1's application environment included the IPM and the visualizations it generated; Group 2's did not. Both groups started with the same document metadata. Participants trained prior to performing the study task and were able to ask questions about the system and the task. At the conclusion of the triage task, they were asked to assess the relevance of each of the 40 documents and were given a short questionnaire about the triage software.


VKB userstudy40-0506.zvkb
File Edit Format Tool Help
Vkb
Classic Obj
Classic
Times New Roman
Aa
12
B
I
Yahoo Search < antimatter >
VKB userstudy-clasic40-0512.zvkb
File Edit Format Tool Help
Vkb
Classic Obj
Classic
Times New Roman
Aa
12
B
I
Yahoo Search < antimatter >

Antimatter - Wikipedia, the free e
http://en.wikipedia.org/wiki/Antim

SPACE.com -- The Reality of Anti
http://www.space.com/scienceas

Home
http://www.matter-antimatte

Antimatter:Mirror of the Universe
http://livefromcern.web.cern.ch/liv

antimatter -- Britannica Online Er
http://www.britannica.com/eb/arti

Science News for Kids: Fea
http://www.sciencenewsfork

Antimatter rocket - Wikipedia, the
http://en.wikipedia.org/wiki/Antim

Antimatter and chemistry
http://www.apsidium.com/theory/

Antimatter
http://www.lbl.gov/abc/Antim

MySpace.com - Antimatter - Acou
http://www.myspace.com/antima

Antimatter - Memory Alpha, the S
http://memory-alpha.org/en/wiki/

Howstuffworks "How Antima
http://science.howstuffworks

Antimatter Everyday

Secrets"Bizarre Facts From Ange

What's the Antimatter?

Antimatter - Wikipedia, the free e
http://en.wikipedia.org/wiki/Antim
... particle physics and quantum chemistry, antimatter is the extension of the ... Furthermore, mixing of matter and antimatter would lead to the annihilation of ...

SPACE.com -- The Reality of Anti
http://www.space.com/scienceas
Antimatter is created and annihilated in stars every day. ... Wild Posters Celebrate 100th Birthday of Dour Father of Antimatter ...

Home
http://www.matter-antimatter
Updated Periodic Table con
s 218 Matter-AntiMatter elem
s. ... Scientists have discove
hat comets are natural sour
of antimatter. ...

Antimatter:Mirror of the Universe
http://livefromcern.web.cern.ch/liv
Discover what antimatter is, where it is made, and how it is already a part of our lives.

antimatter -- Britannica Online Er
http://www.britannica.com/eb/arti
... online encyclopedia article on antimatter: substance composed of subatomic ... The antimatter particles corresponding to electrons, protons, ... antimatter ...

Science News for Kids: Fea
http://www.sciencenewsfork
Scientists are creating parti
of antimatter to learn why th
verse has so few. ... using h
ech machines to make parti
of so-called antimatter. ...

Antimatter rocket - Wikipedia, the
http://en.wikipedia.org/wiki/Antim
... products of antimatter annihilation for propulsion, those that heat a working ... 4 Difficulties with antimatter rockets. 5 See also. 6 References ...

Antimatter and chemistry
http://www.apsidium.com/theory/
The topic antimatter and chemistry, it appears on the first gaze unusually, ... a chemical contemplation of the antimatter is indirectly possible over the very ...

Antimatter
http://www.lbl.gov/abc/Antim
... from a distance matter an
timatter would look essentia
dentical, there ... on the low
erved abundance of antimat
n the cosmic rays, which are

MySpace.com - Antimatter - Acou
http://www.myspace.com/antima
MySpace music profile for Antimatter with tour dates, songs, videos, pictures, blogs, band ... Moss' performance on ANTIMATTER's fourth album, "Leaving ...

Antimatter - Memory Alpha, the S
http://memory-alpha.org/en/wiki/
Antimatter was a generalized term used to describe a state of matter which was ... harmonics to match the antimatter containment effectively severing the Koinonian

Howstuffworks "How Antima
http://science.howstuffworks
Antimatter has the ability to
incredible amounts of ener
a very small ... Antimatter is
ctly what you might think it is
e opposite of normal ...

Antimatter Everyday
http://livefromcern.web.cern.ch/liv

Secrets"Bizarre Facts From Ange
http://www.danbrown.com/secret

What's the Antimatter?
http://www.cem.msu.edu/~c

**Figure 7. Initial layout for Group 1 (left) and Group 2 (right).**

## 8.2 Final Organizations

In both study conditions, participants were able to use any of the basic VKB functionality they found helpful and were free to examine the documents in Firefox. The basic VKB functionality gave participants the ability to change visual attributes of document objects (background color, border color, border width, and size), use links, add textual comments to their workspaces, and create sub-collections.

As we have observed in previous studies, participants went about the triage task in significantly different ways regardless of their group. Some participants began by reading a few documents carefully to get a sense of the topic and the types of documents they were working with, and then proceeded to organize the remaining documents quickly. Some participants began by organizing documents based on visible text, and then refined their initial organization as they read further. Other participants arranged the documents in VKB's workspace using both space and color; others used space alone, stacking documents or creating sub-collections according to perceived categories. A third set of participants organized documents using the objects' visual attributes without moving them. Participants in Group 1 perceived annotations as one way of going about the task and of communicating their intent to the teacher (the hypothetical consumer of the results). Seven participants in this condition actively highlighted or added notes.

## 8.3 Activity Data and Analysis

While users were performing the task, user actions in VKB and the Web browser were logged. The log of Web browser activity included time spent on a document, mouse clicks, scrolls, focus-in, focus-out, mouse over, mouse out, highlight, and note. VKB logged document organizing activity including changes to objects' spatial or visual attributes and objects or collection creation. The data analysis focused on evaluating whether the interest-based visualizations helped users perform the triage task from three perspectives that were sources of difficulty for participants in prior studies: (1) keeping track of progress; (2) identifying documents of interest; and (3) reducing the conflict between user-authored and system-computed visualizations. We use these three evaluation goals to organize our results.

### *8.3.1 Keeping Track of Progress: Task Switching*

Our previous document triage study showed that users had difficulty keeping track of progress as they shifted their attention quickly between the triage-related applications or among multiple documents. One aim of this study was to evaluate whether VKB 3's new thumbnail-based document surrogates make objects more recognizable (and potentially more useful) without opening them.

The average session reading times for the two groups were compared. For the purposes of this study, a session is defined by a continuous series of logged interactions that refer to the same document. That is, a user may read, scroll, annotate, scroll some more, and continue reading; this sequence of connected actions is considered to be a session. Group 1's average session reading time (10.69 seconds) is almost 2.5 times (246%) greater than Group 2's (4.34 seconds). Group 1's total number of sessions (1525) is considerable fewer than Group 2's (2148) even though Group 1's total reading time (15,338,798 seconds) is almost double that of Group 2 (8,636,299 seconds). These results indicate that Group 1 looked at fewer documents, but when they actually engaged with a document, they spent far longer doing it.

Figure 8 shows the distribution of the participants' session reading times across the two groups. The histogram shows that Group 2's members had more instances of relatively short reading sessions than Group 1's members, and Group 1's members had more instances of relatively long reading sessions. Based on the Mann-Whitney test, participants in Group 1 significantly differ in session reading time from those in group 2 (U=1391140, $p < 0.0001$). It is notable that Group 2 participants had substantially more episodes in which they focused on a document very briefly; it seems that VKB 3's facility for displaying document thumbnails helped alleviate the phenomenon of needing to put a document into focus just to remind oneself of what it is.

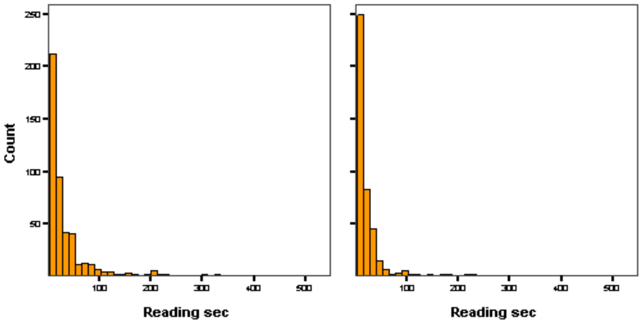


**Figure 8. Session reading time of group 1 (left) and 2 (right)**

The average number of times each document was opened from VKB (by double-clicking on the document's surrogate) was also compared; this average was computed by dividing the total number of unique documents each participant opened during the triage task by the number of document-opening events. This average should reveal how many times documents were re-opened. This average for participants in Group 1 (1.25) is 14% lower than the average for Group 2 (1.46), even though Group 1's average reading time (25.55 seconds) is higher than Group 2's (14.40 seconds). This comparison between Groups 1 and 2 indicates that Group 1 re-opened fewer documents than Group 2 did, and they spent more time reading the ones that they did open. Based on the Mann-Whitney test, participants in Group 1 significantly differ in how many documents they re-opened from participants in Group 2 (U = 55828, p = 0.010).

These results show that Group 1 participants switched their attention between triage-related applications and among individual documents significantly less than Group 2 participants did. This improvement in maintaining focused attention implies that the new visualizations were effective in helping participants keep track of their progress through the documents during the task. They found it less necessary to re-open documents and they were able to spend longer reading individual documents.

### *8.3.2 Documents of Interest: Time on Documents*

In our previous document triage study, users often spent a substantial amount of time examining documents that they later decided were not useful. This finding suggests that users may benefit from system assistance in locating interesting documents for the task. Did the new visualization help participants find the documents they needed? If it did, participants in Group 1 should have spent more time reading and manipulating the documents they later assessed as relevant than did participants in Group 2.

Table 3 shows the relationship between reading time and each participant's evaluation of document relevance based on

Spearman's correlation coefficient. All but one participant had a positive correlation, indicating they spent more time on documents they evaluated positively. The values for participants exhibiting a statistically significant correlation ($p < 0.05$) are shaded. For six out of ten Group 1 participants, the correlation is significant; the correlation is significant for only two out of ten Group 2 participants. The higher degree of correlation implies that Group 1 participants generally spent more time on relevant documents than Group 2 participants did; thus we may infer that the new visualizations in the enhanced applications were effective in helping participants find documents of interest.

**Table 3. Correlation of reading time and document relevance.**

| Group 1 | | | Group 2 | | |
|---|---|---|---|---|---|
| **ID** | **Coef.** | **Sigma** | **ID** | **Coef.** | **Sigma** |
| 1 | 0.429 | 0.018 | 11 | 0.277 | 0.093 |
| 2 | 0.397 | 0.014 | 12 | 0.111 | 0.565 |
| 3 | 0.356 | 0.087 | 13 | 0.210 | 0.205 |
| 4 | 0.409 | 0.011 | 14 | - 0.148 | 0.376 |
| 5 | 0.576 | 0.008 | 15 | 0.367 | 0.024 |
| 6 | 0.206 | 0.214 | 16 | 0.633 | < 0.0001 |
| 7 | 0.137 | 0.412 | 17 | 0.116 | 0.489 |
| 8 | 0.438 | 0.006 | 18 | 0.114 | 0.495 |
| 9 | 0.629 | < 0.0001 | 19 | 0.101 | 0.547 |
| 10 | 0.170 | 0.309 | 20 | 0.240 | 0.147 |

### *8.3.3 Reducing Visual Conflict: Expression in Color*

In our past research, users expressed their interpretation of documents during the triage task by changing the visual attributes of VKB's document surrogates. However, in an effort to bring certain documents to the users' attention, in a previous design, VKB also changed objects' visual attributes. This overloading of the objects' visual characteristics caused a conflict between user-authored and system-generated visualizations: many participants in the previous study were unwilling to overwrite the system's visualizations and became passive in changing objects' visual attributes to express their own interpretations. The new visualizations try to reduce this source of conflict. To evaluate whether this approach was effective, we counted the frequency with which participants in each group changed the background color of VKB's document surrogates. The average number of color-changing events for Group 1 participants is 48% higher than for Group 2 participants. This difference, while not significant based on the Mann-Whitney test (U=0.250, p = 0.250), indicates that participants using the enhanced applications were not dissuaded from visual expression by the visualization.

## 8.4 Questionnaire Results

In post-task questionnaires, eight out of ten Group 1 participants answered that the within-document visualizations were helpful in finding new information of interest; two participants responded neutrally. Three out of ten participants answered that the within-document visualizations distracted them from their reading; one participant answered neutrally; and the remaining six participants said that the new within-document visualizations were not distracting. Thus we can conclude that most of the participants found the new within-document visualization technique to be a helpful means of finding new information of interest in the document they were reading, but some of them found that the visualizations were also distracting.

Nine out of ten participants responded that the enhanced visualization capabilities in VKB were helpful in identifying documents of interest within the workspace. Only one participant answered neutrally. Two out of the ten participants in Group 1 answered that VKB's computed visualizations distracted them from their reading (1 neutral, 6 disagrees, and 1 strongly disagrees). All of the participants found that VKB's visualizations were helpful in organizing documents. Two out ten participants answered that the computed visualizations distracted them when they were trying to organize documents (2 neutral, 3 disagrees, and 3 strongly disagrees). In summary, most of the participants found that the new visualization capabilities in VKB were helpful in finding (identifying) documents of interest and in organizing documents even though a few participants found the computed visualizations to be distracting.

## 9. CONCLUSIONS AND FUTURE WORK

We have developed an extensible architecture that supports triage across multiple applications. The central element of this architecture is the Interest Profile Manager, which receives information about user activity from the individual applications and broadcasts inferred user interests back to the applications. The IPM consolidates functionality necessary to characterize user interests, including the ability to collect, parse, and determine similarity among common forms of Web documents. In our example instantiation of this architecture, the IPM communicates with a Visual Knowledge Builder workspace and an annotation-enabled Web browser. The examples used in Section 7 show how the user's actions in either application can generate assistive visualizations in both applications.

To extend this infrastructure with additional applications, the applications must be able to record and aggregate user activity and communicate it to the IPM and/or receive and use broadcasts from the IPM. Applications need not do both; it may make sense for an application that incorporates a non-interactive visualization technique to receive information about inferred user interests without sending any information to the IPM about user activity. Similarly, an application may be interactive, and may offer considerable insight into a user's interests, but it may not make sense to modify anything in that application accordingly; for example, if the user is writing a paper while she is performing triage, the topics that emerge in the paper may be a very effective source of interest profile data.

The classification of documents of into different user interests in the current IPM is based solely on explicit user expression in a single application. For example, documents or subdocuments that a user colors red will generate red visualizations for documents or subdocuments the IPM analyzes and finds to be similar. Other applications have identified classifications of documents by clustering the documents based on textual analysis or image processing [9]. Such capabilities may help determine when the user has multiple interests that are expressed using the same color or when the user has used different colors to express the same interest.

To make the IPM more readily extensible, the IPM needs to incorporate an abstract model that characterizes the expressive and presentational capabilities of applications. For example, such a model would specify that VKB allows users to assign colors to document surrogates to informally express their interest in a document, their understanding of what a document is about, or a general assessment of its worth to them. By contrast, WebAnnotate displays documents' contents; thus any user expression of interest or other interpretation conveyed through annotations happens at a sub-document level. Components of such an application model include the granularity of information presented, persistent forms of user expression, transient forms of user interaction, and visualization methods supported.

We have evaluated the effectiveness of the visualizations (including the enhanced presentation of document surrogates) and have found that they are successful in allowing people to do less switching among documents and applications; in promoting longer engagements with individual documents; and in recommending interesting new documents and passages within documents based on what the user has indicated interest in already. We also found that users were relatively comfortable with the new capabilities and only a few of them found the computed visualizations distracting. Although the study was not designed to test the IPM architecture or basic capabilities of the applications, we found that they performed well during the study.

Many types of information work require more than one application; triage is a good example of this phenomenon. Thus we expect our results—evidence of interest from one application may be abstracted and combined with evidence of interest from other applications and re-formulated to assist users—to apply to other suites of applications and other types of information work.

**Acknowledgments**. This work was supported in part by National Science Foundation grants DUE-0938074 and DUE-0840715.

## 10. REFERENCES

[1] Badi, R., Bae, S., Moore, J.M., Meintanis, K., Zacchi, A., Hsieh, H., Shipman F., and Marshall, C.C., Recognizing user interest and document value from reading and organizing activities in document triage. *Proc. of IUI*, 2006, 218-225.

[2] Bae, S., Badi, R., Meintanis, K., Moore, J.M., Zacchi, A., Hsieh, H., Marshall, C.C., and Shipman, F.M. Effects of display configurations on document triage. *Proc. of INTERACT*, 2005, 130-143.

[3] Bates, M., The design of browsing and berrypicking techniques for the online search interface, *Online Review*, 13(5), 1989, 407-424.

[4] Buchanan, G. & Loizides, F. Investigating document triage on paper and electronic media, *Proc. ECDL*, 2007, 416-427.

[5] Card, S. & Nation, D. Degree-of-interest trees: a component of an attention-reactive user interface. *Proc. of AVI*, 2002.

[6] Cantador, I. and Castells, P. Multilayered semantic social network modeling by ontology-based user profiles clustering. *Proc. of CIKM*, 2006, 334-349.

[7] Claypool, M., Le, P., Waseda, M., and Brown, D. Implicit interest indicators, *Proc. of IUI*, 2001, 33-40.

[8] Cool, C., Belkin, N.J., Kantor, P.B. Characteristics of text affecting relevance judgments. *Proc. of National Online Meeting*, 1993, 77–84.

[9] Cutting, D., Karger, D., Pedersen, J. and Tukey, J.W. Scatter/Gather: a cluster-based approach to browsing large document collections, *Proc. SIGIR*, 1992.

[10] Czerwinski, M., Dumais, S., Robertson, G., Dziadosz, S., Tiernan, S. & van Dantzich, M. Visualizing implicit queries for information management and retrieval. *Proc. CHI*, 1999.

[11] d'Entremont, T., and M. A. Storey, Using a degree-of-interest model for adaptive visualizations in Protégé, *Proc. of International Protégé Conference*, 2006.

[12] Farzan, R. and Brusilovsky, P. Social navigation support through annotation-based group modeling. *Proc. of User Modeling*, 2005.

[13] Garrigós I., and Gómez J. Modeling user behaviour aware websites with PRML. *Proc. of WISM*, 2006.

[14] Grudin, J., Groupware and social dynamics: eight challenges for developers, *CACM*, 35:92 – 105, 1994.

[15] Gunduz, S., & Ozsu, M. A user interest model for web page navigation. *Proc. of DMAK*, 2003, 46-57.

[16] Hearst, M. TextTiling: segmenting text into multi-paragraph subtopic passages, *Computational Linguistics*, 23(1), 33-64.

[17] Kim, S. & Fox, E.A. Interest-based user grouping model for collaborative filtering in digital libraries. *Proc. of ADL*, 2004

[18] Marshall, C.C. and Shipman, F. Spatial hypertext: designing for change", *CACM*, 38(8), 1995, 88-97.

[19] Marshall, C.C. and Shipman, F. Effects of hypertext technology on the practice of information triage. *Proc. of HT*, 1997, 124-133.

[20] Nichols, D., Pemberton, D., Dalhoumi, S., Larouk, O., Belisle, C., Twidale, M. DEBORA: developing an interface to support collaboration in a digital library. *Proc. of ECDL*, 2000, 239-248.

[21] Qiu, F. and Cho, J. Automatic identification of user interest for personalized search. *Proc. of WWW*, 2006.

[22] Qu, Y., and Furnas, G.W., Sources of structure in sensemaking. *CHI '05 Extended Abstracts*, 2005

[23] Renda, M. E. and Straccia, U. A personalized collaborative digital library environment: a model and an application. *Information Process Management* 41(1), 2005, 5-21.

[24] Sarwar, B., Konstan, J., Borchers, A., Herlocker, J., Miller, B., and Reidl, J., Using filtering agents to improve prediction quality in the GroupLens research collaborative filtering system. *Proc. of CSCW*, 1998.

[25] Schilit, B.N., Golovchinsky, G., and Price, M.N. Beyond paper: supporting active reading with free form digital ink annotations, *Proc. of CHI*, 1998, 249-256.

[26] Shipman, F., Moore, J.M., Maloor, P., Hsieh, H., and Akkapeddi, R. Semantics happen: knowledge building in spatial hypertext, *Proc. of HT*, 2002, 25-34.

[27] Shipman, F., Hsieh, H., Airhart, R., Maloor, P., and Moore, J.M. The Visual Knowledge Builder: a second generation spatial hypertext, *Proc. of HT*, 2001, 113-122.

[28] Shipman F., Hsieh, H., Moore, J.M., & Zacchi, A. Supporting Personal Collections across digital libraries in spatial hypertext. *Proc. of JCDL*, 2004, 358-367.

[29] Shipman, F., Price, M., Marshall, C.C., & Golovchinsky, G. (2003). Identifying useful passages in documents based on annotation patterns, *Proc. of ECDL*, 2003, 101-112.