Recognizing User Interest and Document Value from Reading and Organizing Activities in Document Triage

Rajiv Badi, Soonil Bae, J. Michael Moore, Konstantinos Meintanis, Anna Zacchi, Haowei Hsieh, Frank Shipman
Department of Computer Science
Texas A&M University
College Station, TX 77843-3112 USA
1 979 845 5534
shipman@csdl.tamu.edu

Catherine C. Marshall
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052 USA
cathymar@microsoft.com

ABSTRACT
People frequently must sort through large sets of documents to identify useful materials, for example, when they look through web search results. This document triage process may involve both reading and organizing, possibly using different applications for each activity. Users' interests may be inferred from what they read and how they interact with individual documents; these interests may in turn be used as a basis for identifying other documents or document elements of potential interest within the set. To most effectively identify related documents of interest, activity data must be collected from all applications used in document triage. In this paper we present a common framework (the Interest Profile Manager) for collecting and analyzing user interest. We also present models for detecting user interest based on reading activity alone, on organizing activity alone, and on combined reading and organizing activity. A study comparing document value calculated using the different models shows that incorporating interest information from both reading and organizing activity better predicted users' valuation of documents. This difference was statistically significant when compared to using reading activity alone.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *search process, selection process.*

General Terms
Design, Experimentation, Human Factors.

Keywords
Document Triage, Information Triage, Sensemaking, User Interest Recognition, User Interest Modeling, Visual Knowledge Builder.

1. INTRODUCTION
With the growth of the Internet and other information sources, locating material about a topic of interest is less of a problem than determining which documents are the most valuable for the task at hand. Frequently the desired information is obscured within a long list of potentially relevant resources. People are inundated with so much information that they spend the majority of their time sifting through documents rather than focusing on the task itself. New techniques can improve users' ability to cull the best documents from a large set to gather the information they need.

Document triage is the practice of quickly determining the usefulness and relevance of documents in a collection of documents (e.g. from a search engine). In triage, the attention of the user shifts from document to document to contextual overview (e.g. a list of search results, bookmarks, or a desktop or visual overview of documents). Hence, document triage involves extensive reading (engagement with multiple documents at once) and hyper-extensive reading (looking at document components and fragmentary information), as opposed to intensive reading (deep reading of a single document).

As people sort through a set of documents, they begin to pull out and organize those documents they find most relevant. They bookmark websites, place papers into piles according to topic, or even create categories and lists. Alternatively, documents may be disposed of or ignored. With extremely large sets of documents, people will frequently stop sorting when what they have is perceived as good enough. Consequently, more useful materials may be left unread.

As people engage in document triage, their activities may indicate their interests and needs more effectively then the query they initially formulated. Tools that interpret these activities and infer interest can provide cues to identify documents that are more likely to be of interest as the task proceeds.

1.1 Supporting Document Triage
We are exploring ways to support people performing triage tasks based on inferred user interest. The requisite processing can be broken into four steps: (1) recognizing user interest and document value; (2) representing user interest; (3) identifying other documents of potential interest; and (4) notifying users – possibly through visualizations – that there are documents that match these inferred interests.

In describing these four steps, we will refer to three types of applications involved in document triage: *document overview software* presents results from explicit and implicit searches, *document reading software* displays the content of a document, and *document organization software* records the results of the triage and sensemaking activity.

1.1.1 Recognizing User Interest & Document Value

People demonstrate their interests either explicitly by rating the documents they read or implicitly through their interactions with the documents. One drawback of explicit ratings is that people tend to read many more articles than they rate [10]. On the other hand, recognizing implicit user interest requires software-specific techniques: implicit expression of document interest in overview software is different than implicit expression of document interest in reading software.

1.1.2 Representing User Interest

Multiple software applications are often used during a triage task. Consider someone who is preparing a presentation on the effects of rising oil prices. First a web browser may be used to do a search. The results of this search may be viewed in a variety of reading applications including a web browser, Acrobat Reader, media players, etc. Finally, links to useful resources or excerpts from them may be copied into a word processor. Each application can have its own methods for recognizing user interest, but the representation of inferred interest must be shared across all applications for this to be useful to the entire triage process. User interest can be represented in terms of a set of documents or document components, abstractions of these elements (e.g. term vectors and metadata), or actions on a document (e.g. scrolling patterns and time spent viewing a document).

1.1.3 Recognizing Documents of Potential Interest

As applications receive interest information from their own use and the use of other applications, they can begin to identify entire documents related to this interest (e.g. among search results in overview software) or document portions that match the interest (in reading software). The method for establishing whether a document or document component is related to a recognized interest depends on how the interest is represented.

1.1.4 Visualizing Interest Information

Once information relevant to user interest is identified, users must be notified. There are a variety of approaches for notifying users, including suggestion mechanisms and visualization techniques. For example, in overview software, relevant documents could be identified by visually distinguishing proposed documents of interest. In a reading application, sections of relevant text within the document could be highlighted with a yellow background.

This paper addresses the first step, recognition of user interest. Specifically, the focus is on recognizing user interest based on a user's behavior across multiple applications involved in triage activity. In the next section, we discuss approaches for recognizing user interest. That is followed by an introduction to our system for facilitating the sharing of interest information across applications. We then describe four models for identifying user interest and compare these models against user-specified document values.

2. RECOGNIZING USER INTEREST

Much research has gone into gathering and recognizing user interest. Systems developed to recognize user interest may employ implicit indicators, explicit indicators, or a combination of both.

2.1 Explicit Interest Indicators

With explicit indicators (e.g. ratings), users tell the system how interesting or uninteresting a given document is. Explicit indicators are well-understood, easy to implement, and fairly precise. However, stopping to enter explicit ratings can interrupt normal patterns of browsing and reading [3] and may impose an increased cognitive load on the user. Users may stop rating items when they perceive that there is no benefit in providing these explicit ratings [5]. Studies of GroupLens found that users assigned explicit ratings to many fewer documents than they read [10]. Thus, although explicit ratings are a fairly precise expression of interest, their efficacy is limited.

2.2 Implicit Interest Indicators

Implicit indicators are a viable alternative to explicit ratings. Nichols [9] identifies some implicit interest indicators and discusses the potential of implicit ratings. He concludes that "the limited evidence available suggests that implicit ratings have great potential, but their effectiveness remains unproven."

Indeed, determining what characteristics to use to infer user interest is difficult and context dependent. For reading applications, time spent reading a document has been found to be a good indicator of interest in many situations [2, 3, 7, 8]. Conversely, Kelly and Belkin [6] find no general, direct relationship between display time and usefulness. They note that display times varied according the specific task and user – emphasizing the contextual nature of recognizing interest. Identifiable types of user interactions such as scrolling and mouse events have been found to be predictive as well [3, 4], although these studies conflict on the correlation between mouse movement and user interest.

User annotations are another type of implicit indicator considered in the literature. Studies of law students showed that their annotations may be used to predict interest in document passages, as evidenced by their later citations [13]. Analysis demonstrated that some types of annotations (for example, interpretive comments) demonstrated greater user interest than others (for example, highlights), and that passages that were annotated in several different ways were of the greatest interest.

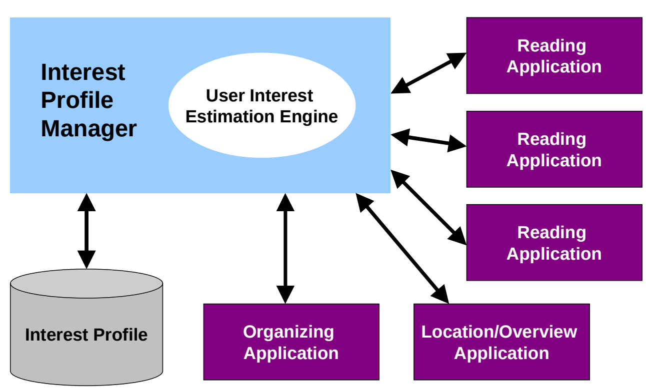

Figure 1: Interest Profile Manager

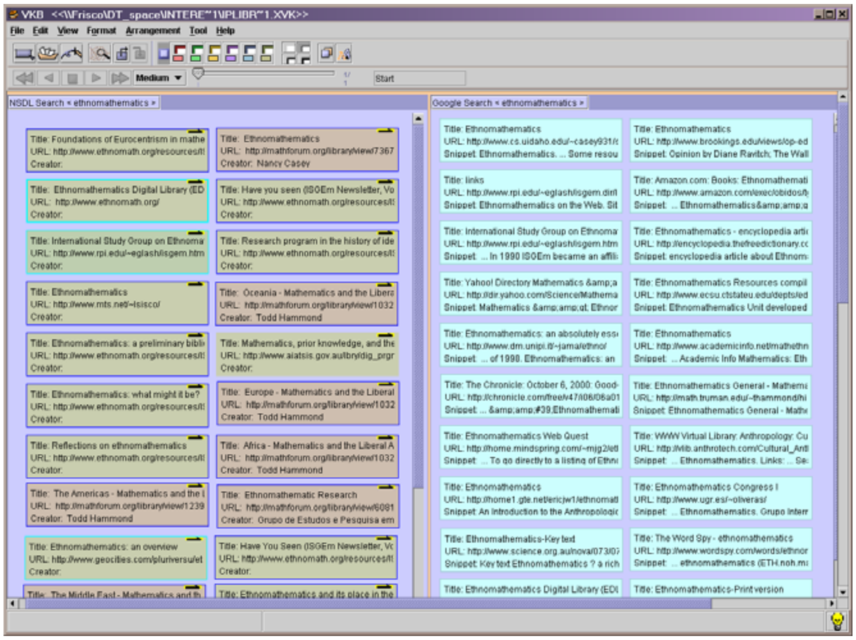

Figure 2: VKB workspace with search results

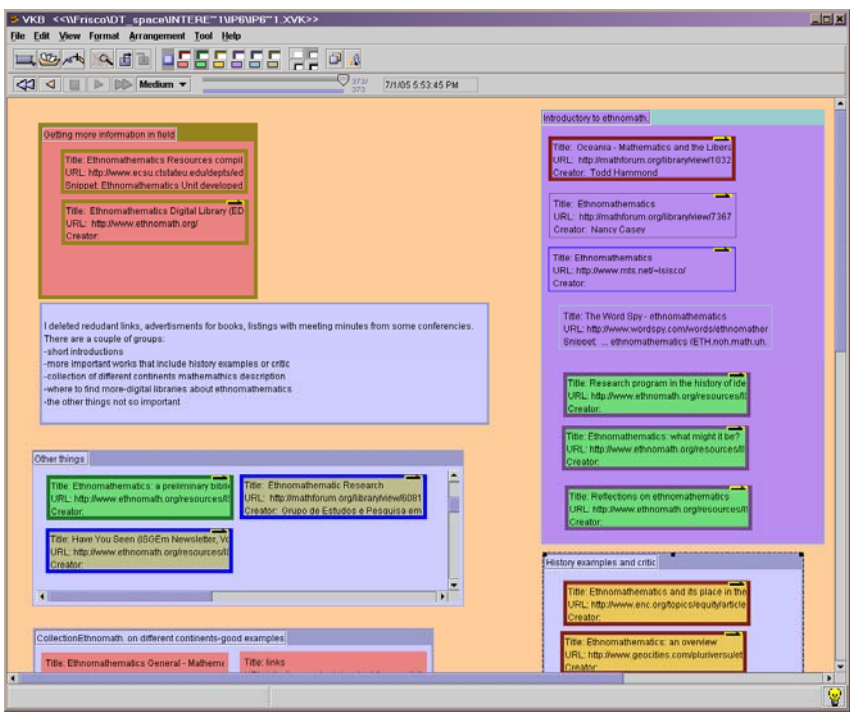

Figure 3: VKB workspace after organizing activity

3. INTEREST PROFILE MANAGER

Prior work has focused on a single reading application as the source for interest indicators. However, as indicated earlier, triage occurs in the context of multiple applications. We have created the Interest Profile Manager as the basis for determining, sharing and storing user interest based on interest indicators from multiple applications (Figure 1). The Interest Profile Manager acts as an independent server. Applications connect to the server and store and retrieve interest information via a linkable software library. The library can be used to modify existing programs to interface with the Interest Profile Manager or to create connector applications that interface the Interest Profile Manager to an existing application with its own interface library such as Internet Explorer. In this way overview, reading, and organizing applications can connect to the Interest Profile Manager.

As users work, client applications send interest-related activity information to the Interest Profile Manager as a set of attributes and values; this information is stored in the Interest Profile. Changes to the Interest Profile are propagated to all clients subscribing to such updates. The Interest Profile Manager includes a User Interest Estimation Engine that evaluates user interest based on information in the Interest Profile and its interest models. The interest profile is augmented with the results calculated from the interest models, and these calculations are communicated to all subscribed clients.

For our initial implementation, we modified the Visual Knowledge Builder (VKB) [11] to act as both the overview application and the organizing application. In VKB, search results are returned as information objects inside of a two-dimensional space called a *collection* (Figure 2). Each information object refers to a search result and double clicking on an object opens the corresponding URL. VKB also allows users to add text to information objects and associate metadata with both collections and objects. Moreover, visual attributes including border width, object size, and background and border color can be edited for objects and collections. The addition and manipulation of text, metadata, and visual attributes act as implicit interest indicators. Figure 3 shows the workspace after organizing activity.

While we plan to integrate multiple reading applications using the Interest Profile Manager, our initial implementation was limited to an instrumented version of Microsoft's Internet Explorer (IE). IE was augmented to gather and log user activity and to communicate with the Interest Profile Manager. The next section describes how user interest models were developed based on use of IE and VKB.

4. INTEREST MODELS

Our interest models currently rely on implicit indicators. This information for inferring interest is gathered unobtrusively from a reading interface (currently IE) and an organizing interface (currently VKB). Data gathered includes document attributes, data coming from the document reading activity, and data coming from the document organizing activity (Table 1).

Table 1: Data collected

Category	Parameter
Document Attributes	Number of characters (words/page)
	Number of links
	Number of images
	File size of a document (Bytes)
Document Reading Activity (IE)	Time spent in a document
	Number of mouse clicks
	Number of text selections
	Number of scrolls
	Number of scrolling direction changes
	Time spent scrolling
	Scroll offset
	Total number of scroll groups
	Number of document accesses
Document Organizing Activity (VKB)	Number of object creations
	Number of object moves
	Number of object resizes
	Number of object deletions
	Number of content changes
	Number of background color changes
	Number of border color changes
	Number of border width changes
	Number of font changes
	Number of font color changes
	Number of canvas color changes
	Number of changes in z-order
	Number of transparency changes

Document attributes are characteristics inherent to the documents themselves, independent of users' interactions with them. We primarily consider document length (e.g. number of pages, number of characters and number of words), number of links and images in a document, and file size.

Document reading activity includes user actions during passive reading in the reading application. This consists of time spent viewing a document, number of mouse clicks, number of text selections, characteristics of the user's scrolling behavior, and frequency of document access.

Document organizing activity refers to user actions on objects representing documents in VKB. Creating collections, placing document objects in collections, and changing spatial or visual attributes of categories and document objects are included in this activity [12]. Through these actions, users may express their interpretation of the documents and their interrelations.

A prior study [1] showed a correlation between user interest with user events such as reading time and a few document attributes. A more comprehensive correlation analysis revealed additional user events for estimating document value from organizing activity, reading activity, and document attributes. Table 2 shows attributes correlated to explicit user-assigned ratings. With one exception, all p-values are less than 0.005. "Number of characters" has a p-value of 0.01 but is included because interviews with study subjects revealed that they consider the amount of content in a document to be an important criterion for classifying documents. The only negative correlation, "Number of object deletions" makes sense since removing a document should correspond to little or no interest in the document. As has been shown in prior studies, time spent viewing a document positively correlates to user interest. Note that four different types of scrolling events correlate to user interest.

Table 2: Results from correlation analysis

Parameter	Pearson coefficient	p-value
Number of characters	0.431	0.010
Time spent in a document	0.527	0.001
Number of scrolls	0.630	<0.001
Scroll offset	0.641	<0.001
Number of scrolling direction changes	0.589	<0.001
Total number of scroll groups	0.590	<0.001
Number of document accesses	0.476	0.004
Number of object moves	0.711	<0.001
Number of object resizes	0.622	<0.001
Number of object deletions	-0.495	0.003
Number of background color changes	0.597	<0.001
Number of border color changes	0.628	<0.001
Number of border width changes	0.525	0.001

The correlation analysis provides a set of parameters for estimating users' interest from their activities in IE (reading application), VKB (organizing application), and document attributes. These are the basis for building mathematical and hand-tuned models to estimate user interest. Since there are significant correlations from both the reading and organizing applications, models of user interest can take advantage of events from both types of application.

4.1 Mathematical Models

Three mathematically-derived models are presented that calculate user interest. Data gathered from the prior study [1] was used to derive these models. This data includes the parameters identified in Table 2 and explicit user assessment of documents. In this case, the explicit user assessment involved each study subject selecting the five most interesting documents and the five least interesting documents out of the corpus of 34 documents after performing a prescribed document triage task. To generate a quantitative characterization of interest, documents rated as being of least interest were assigned an interest value of 0, documents rated as being of high interest were assigned an interest value of 2, and the remainder were given an interest value of 1.

Rather than classifying documents as being of low, average, or high interest, all three mathematical models were developed to produce floating point values between 0 (low interest) and 2 (high interest). A value below 1 indicates the document is of less than average interest while a value greater than 1 indicates a greater than average interest.

The first model is limited to reading activity in IE and document attributes. The second model is limited to organizing activity in VKB. The final model combines reading activity in IE, organizing activity in VKB, and document attributes to estimate user interest.

Table 3: Factor score coefficients for the reading activity model

Parameter	Factor1	Factor2
Number of characters	-0.199	0.537
Number of scrolls	0.123	0.187
Number of document accesses	0.294	-0.229
Time spent in document	0.265	-0.069
Scroll offset	-0.072	0.442
Number of scrolling direction changes	0.227	0.010
Total number of scroll groups	0.258	-0.050

Standardized parameter =
(parameter – mean) / standard deviation.

Factor1 =
– 0.199 * (standardized number of characters)
+ 0.123 * (standardized number of scrolls)
+ 0.294 * (standardized number of visits)
+ 0.265 * (standardize time spent in document)
– 0.072 * (standardized scroll offset)
+ 0.227 * (standardized number of scrolling direction changes)
+ 0.258 * (standardized total number of scroll groups)

Factor2 =
0.537 * (standardized number of characters)
+ 0.185 * (standardized number of scrolls)
– 0.229 * (standardized number of visits)
– 0.069 * (standardized time spent in IE)
+ 0.442 * (standardized scroll offset)
+ 0.010 * (standardized number of scrolling direction changes)
– 0.050 * (standardized total number of scroll groups)

Figure 4: Reading-Activity equations to calculate factors

All three models were developed using aggregated user activity and averaged user evaluation of the documents' value.

4.1.1 Reading-Activity Model

The reading-activity model uses document attributes such as number of characters, images, and links and user events recorded during reading, such as number of scroll events and time spent viewing the document. Correlation analysis between these document attributes/user events and the explicit user ratings on documents (document score) identified seven variables. Many of these variables are correlated with each other, and regression modeling requires independent variables. Factor analysis produced two independent factors based on the seven variables. These factors can be calculated from the factor coefficients generated for each parameter as shown in Table 3. Figure 4 demonstrates how factors are calculated. The multiple regression model based on those factors is:

Document Score = 0.877 + 0.133 * *Factor1*
+ 0.120 * *Factor2*

4.1.2 Organizing-Activity Model

The organizing-activity model uses events recorded during organizing activities such as moving/resizing/deleting objects and changing objects' background or border color. The correlation analysis between these user events and explicit user ratings for documents identified six variables. We have extracted two independent factors through factor analysis. The coefficients for calculating the factors in this model are found in Table 4. Factors are calculated in a similar fashion as Figure 4. The multiple regression model based on those factors is:

Document Score = 0.877 + 0.185 * *Factor1*
– 0.092 * *Factor2*

4.1.3 Combined Model

The combined-activity model is based on document attributes and user events recorded during reading and organizing (i.e. it combines those used in the reading-activity and organizing activity models). In this case, factor analysis identified four independent factors based on the 13 variables. Table 5 shows the factor coefficients for each parameter used in calculating the four factors. The multiple regression model based on those factors is:

Document Score = 0.877 + 0.125 * *Factor1*
+ 0.152 * *Factor2*
+ 0.0662 * *Factor3*
+ 0.0653 * *Factor4*

4.2 Hand-Tuned Model

The previous three models are based on the statistical analysis of prior study data. They were derived so as to minimize the error with respect to the explicit user evaluation of documents in the earlier study. Because the study design required subjects to identify at most five documents as high value and at most five as low value, the remaining 25 considered to be of average value. Thus, the factor analysis for the three mathematical models above resulted in models that are conservative their estimates of how documents vary from average interest.

We defined a fourth hand-tuned model based on a combination of qualitative and quantitative assessment of previous study data. Qualitative data included interviews with subjects about what made documents valuable and analysis of videotapes of subjects as they performed the task. Table 6 summarizes this model. The weight field in the table is the degree to which the parameter contributes to the overall document score. Currently, document attributes constitute 10% of the hand-tuned interest model, reading activity is used for 37.5% of the interest value, and

Table 4: Factor score coefficients for the organizing activity model

Parameter	Factor1	Factor2
Number of object moves	0.331	0.202
Number of object resizes	0.325	0.327
Number of object deletions	0.164	0.820
Number of background color changes	0.210	-0.232
Number of border color changes	0.234	-0.174
Number of border width changes	0.164	-0.105

Table 5: Factor score coefficients for the combined activity model

Parameter	Factor1	Factor2	Factor3	Factor4
Number of characters	-0.143	-0.038	0.549	0.174
Number of scrolls	0.164	-0.142	0.204	-0.091
Number of visits	0.197	0.117	-0.246	0.250
Time spent in a document	0.268	-0.112	-0.057	-0.119
Scroll offset	0.005	-0.140	0.472	0.044
Number of scrolling direction changes	0.244	-0.134	0.035	-0.079
Number of scroll groups	0.266	0.113	0.047	0.137
Number of object moves	0.026	0.291	-0.123	0.149
Number of object resizes	0.066	0.195	-0.070	0.317
Number of object deletions	-0.036	-0.055	0.170	0.826
Number of background color changes	-0.109	0.319	0.040	-0.073
Number of border color changes	0.101	0.339	0.009	0.023
Number of border width changes	-0.091	0.338	-0.139	-0.132

Table 6: Weights for hand-tuned model

Category	Parameter	Weight
Document Attributes (10%)	Number of characters	5
	Number of links	5
Document Reading Activity (37.5%)	Number of scrolls	5
	Number of document accesses	10
	Time spent in a document	10
	Scroll offset	5
	Number of scrolling direction changes	5
	Number of scroll groups	2.5
Document Organizing Activity (52.5%)	Number of object moves	17.5
	Number of object resizes	15
	Number of object deletions	5
	Number of background color changes	7.5
	Number of border color changes	2.5
	Number of border width changes	5

organizing activity forms the remaining 52.5%.

This model was developed to generate a wider range of interest values and to be the starting point for a model that is not tied to one particular document set and user activity, since it is dangerous to develop models from a user study involving a single task and corpus, no matter how many subjects are included. The hand-tuned model will need to be refined as we gather data from follow-on studies of other tasks and document sets.

4.3 Comparison of Models

The capability of the three statistical models to predict user interest on documents has been compared in terms of R^2 and the adjusted R^2 of the models in Table 7. A value of R^2 is a measure of how much of the variability in the outcome of the models is accounted for by the predictors. For the first model, the value is 0.477, which indicates that parameters of reading activity account for 47.7% of the expressed user interest for documents. Similarly, parameters of organizing activity account for 63.6% of user interest. Finally, parameters of both reading and organizing activity account for 70.8% of user interest. The adjusted R^2 shows how the model generalizes. Ideally adjusted R^2 should be the same as, or very close to, the value of R^2.

Table 7:Comparison of mathematical models

Model	R	R^2	Adjusted R^2
Reading-activity model	0.690	0.477	0.444
Organizing-activity model	0.797	0.636	0.613
Combined model	0.841	0.708	0.669

5. MODEL EVALUATION

The comparison of models shows the error is reduced by using a combined model. This reduction is guaranteed due to the increased number of variables available to model the same data. It remains to be seen whether the combination will improve the model's predictions for other users. Thus, a study was conducted to evaluate the effectiveness of the interest models presented in the previous section. The following subsections discuss the study and the results.

5.1 Study

The study took place in the Center for the Study of Digital Libraries at Texas A&M University. Sixteen graduate students and research associates from the university participated. Subject ages ranged from 22 to 32. All subjects had basic familiarity with using a computer and browsing the web and had used computers regularly for five or more years.

Subjects were placed in the role of a research librarian who had to select and organize documents for a high school teacher preparing a class on ethnomathematics, the study of a group's culturally-specific mathematical practices as its members go about their everyday activities. This is the same task and topic as in our prior studies [1, 12].

Since VKB is the application used to organize the documents, subjects were given a brief training on VKB emphasizing features considered relevant for the task, such as the ability to organize information objects (links to websites in this study) in a hierarchy of two-dimensional workspaces. Subjects used the augmented version of Internet Explorer as the reading application.

Subjects were given 20 documents on ethnomathematics from the National Science Digital Library (NSDL) and 20 documents returned from Google. The links to the 40 documents were placed in lists in a VKB workspace, as shown in Figure 2; all subjects received the same links. The documents varied in their level of sophistication, relevance to the task, and length. Although no time limit was set, all subjects performed the task in less than one and a half hours.

Subjects were requested to organize the 40 links for the high school teacher. They determined their own criteria for organizing the links, and were free to add, modify and delete them. They were also free to modify other attributes in the VKB space such as background color, border thickness, font color, etc., as well as creating new text objects or annotations to the VKB space as they deemed necessary.

Following task completion, subjects rated the usefulness of all 40 documents on a scale of 1 (Not Useful) to 5 (Very Useful). Five of the 40 links were not included in the analysis since their activity data was not available (e.g. the augmented IE did not record activity within an embedded viewer such as Adobe's PDF plugin). Subjects were also interviewed to gain additional insight into how they performed the task, their document ratings, and their use of metadata to evaluate document relevance; this interview data could then be triangulated with other data sources.

User actions in VKB and IE were recorded by the Interest Profile Manager for model calculation.

5.2 Results

The 16 subjects' evaluations of each document were averaged and scaled to a continuous value between 0 (least useful) and 2 (most useful). To investigate the accuracy of each model's prediction of user interest, we calculated the residue for each document. The residue is the absolute value of the difference between the explicit user rating and a model's predicted rating. A perfect predictive model would have an average residue of zero. The results shown in Table 8 are averaged across all documents for each model. This shows the overall prediction accuracy for each model. Both models relying on a combination of reading and organizing

Table 8: Residue comparison of models

Model	Average Residue	Standard Deviation
Reading-activity model	0.258	0.192
Organizing-activity model	0.216	0.146
Combined model	0.175	0.138
Hand-tuned Model	0.197	0.134

activity have lower residue values than either of the models based solely on either reading or organizing activity.

Statistical analysis can identify which models produced residues that are significantly different from each other. Based on the results of paired t-test analysis, the residues from the combined model were significantly lower (better) than those for the reading-activity model ($p=0.02$). Also, the residues from the combined model were lower than those of the organizing-activity model ($p=0.07$). The paired t-test analysis did not show a significant difference between the organizing-activity and reading-activity models ($p=0.31$).

This initial version of the hand-tuned model did not perform as well as the combined model. The residues for the hand-tuned model were not found to be significantly different from those of the reading-activity model ($p=0.13$), the combined model ($p=0.37$), or the organizing-activity model ($p=0.56$). The goal of the hand-tuned model is to reduce the influence of the particular task and document set in this and earlier studies. Further studies will be necessary to evaluate whether this is successful.

As previously mentioned, residues greater than zero represent errors in a model's prediction of user interest. Since document interest values are scaled from 0 (low interest) to 2 (high interest), each 0.1 increment in residue represents a 5% error of predicted values from user assessed values. **Error! Reference source not found.** shows a distribution of errors for each model. Ideally, 100% of all errors would fall in the 0-5% range. So, keeping a large proportion of errors towards 0% is desirable. All models have a majority of their errors distributed in the 0-15% range. The combined model has 94% of its error contained within the 0-15% error range. However, both the reading and organizing models have more errors spread beyond 15%. The local peak in the 20-25% range for the organizing model and the peak in the 25-30% range for the reading model indicate a number of these incorrect assessments that could appear random to users.

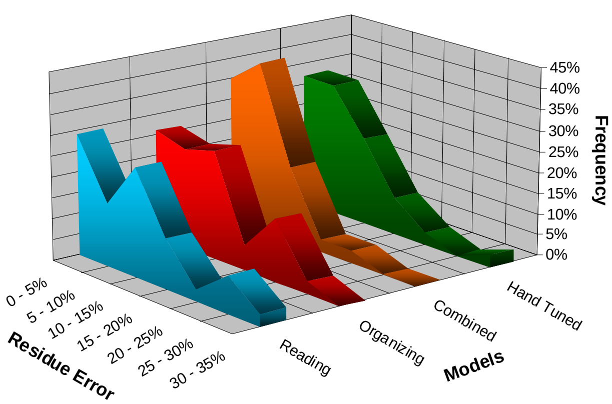

Figure 5: Distribution of residue errors

5.3 Discussion

The results consistently show that the combined activity model performs the best of the three mathematical models while the organizing activity model performs better than the reading activity model. The hand-tuned model performed reasonably well – better than either the reading or organizing activity models.

The prior analyses are based on the aggregated user activity and averages of document valuations. We are currently investigating each model's predictive capability when applied to individual data. Since the mathematical models were developed based on aggregated data, they are not likely to perform well for individuals. On the other hand, the hand-tuned model has been developed with an understanding of how triage practices vary between users.

To understand these differences in individual document triage practices, we looked to the qualitative data gathered in the interviews and the recordings of how the subjects performed the task. In the interviews, subjects were asked why they rated documents lower and higher, about revisiting documents, and about their overall organizational strategy.

Subjects claimed to rate documents mainly based on content. Some subjects considered web pages with more information as useful, e.g. rating academic journals and some comprehensive sources as more useful, while others assigned them low ratings because they felt that the teacher would prefer briefer introductory material for his/her class. Since document length is viewed as positive by some users and negative by others, models of interest should limit the weight they place on document length as a predictor of document value. An adaptive model that attempts to recognize whether the user prefers long or short documents could be beneficial but this is likely to be very task dependent.

Metadata was also used to evaluate documents. Many subjects relied on the domain names to assess a document's authority. Therefore they rated documents from .edu domains and digital libraries higher than other documents. Several subjects may have relied on the update frequency ("Last updated") of the web pages as a criterion for rating them, evidenced by the fact that they rated recently updated documents higher. Some subjects rated a few documents without visiting them, basing their decisions on metadata alone. For example, the recordings reveal that a subject moved an Amazon link to a collection based on its URL and title (visible in the VKB symbol), and subsequently assigned the document a rating of 4 (out of 5). These examples imply that, while not currently part of our models, metadata could be helpful in recognizing user interest.

Subjects revisited documents for different reasons. One subject visited most links cursorily in the first pass "to see what was out there," classified most documents during the second pass, and then classified the remaining documents in the third pass. Another subject revisited some of the documents since he did not recall if he had already visited them. Even so, models of interest can make use of the number of times a document is accessed when combined with other evidence to infer user interest.

Another activity to consider in modeling interest is deletion of references from the overview. In the study, subjects were not given specific instructions about discarding links. Some users chose to delete links that they felt were not useful. Others chose to place such links in separate collections giving those collections labels such as "Other". Some subjects chose not to delete links and left them in their original collections without moving them to any new collections they created. One of the subjects deleted most of the links leaving only ten links in the VKB space. The subject reported that the ten links would suffice for the teacher, and that many of the other links could be reached from them. This subject provided explicit ratings for 34 documents. One of the hypotheses going into the study was that subjects deleted links to documents that were not useful. In this case some useful documents were removed. However, in the context of this task, the remaining documents were relatively more valuable than those removed. In general, deleting documents and leaving documents unaltered is evidence of less interest although there are exceptions.

The use of color in the overview was idiosyncratic based on each subject's strategy for completing the task. One subject used elaborate color coding for web pages; red for good examples, purple for introductory material, and thick red borders for good examples that were already assigned a different color. The subject reported that she would set the color and border thickness of a symbol after reading the corresponding web page and drop the matching symbols into collections in the second pass. Another subject came up with a chapter based classification, using color to show the relationship between the chapter headings and their content. Since recognizing how specific organizational actions should be interpreted is extremely difficult, our models are limited to using the number of visual manipulations instead of trying to interpret specific manipulations.

6. CONCLUSIONS

Document triage is the process of identifying valuable documents within a set of relevant resources. We are investigating methods for actively supporting document triage.

Triage tasks involve culling larger sets of relevant resources by skimming or otherwise reading bits of them, and selecting and organizing desired resources for further use. Thus, the triage process may involve multiple applications, including browsing and overview tools, reading tools, and document organization and management tools.

User activity in these applications often corresponds to user interest in resources and can be the basis for supporting later triage activity. In the past, each application involved in the triage task would have to try to infer the user's interest. The Interest Profile Manager allows applications to communicate information about user activity during the triage task and to estimate user interest in documents based on that activity.

A reading-only activity model, an organizing-only activity model, and a combined activity model were developed based on a prior study of document triage practice. These three models were then compared to explicit user statements of interest in a subsequent study. The evaluation showed that inferred document values calculated using the models incorporating interest information from both reading and organizing activity better predicted users' valuation of documents. This difference was statistically significant when comparing the mathematically-derived combined model to the reading-activity model.

All of these models were developed from aggregated user activity and average document assessments from the earlier study. They are unlikely to perform as well for individuals due to the idiosyncratic nature of document triage practice. By increasing the range of user interactions being used to infer interest, we hope to generate interest models that are more tolerant of individual user and task differences. The hand-tuned model is our first step towards such a model.

7. ACKNOWLEDGEMENTS

This work is partially funded by NSF grants IIS-0438887 and DUE-0226321 and a grant from Microsoft.

8. REFERENCES

[1] Bae, S., Badi, R., Meintanis, K., Moore, J.M., Zacchi, A., Hsieh, H., Marshall, C., Shipman, F. "Effects of Display Configurations on Document Triage," *Proc. of IFIP Interact Conference*, 2005, pp. 130-143.

[2] Chan, P."A Non-Invasive Learning Approach to Building Web User Profiles," *Workshop on Web Usage Analysis and User Profiling*, 1999, pp. 7-12.

[3] Claypool, M., Le, P., Waseda, M., Brown, D. "Implicit Interest Indicators," *Proc. of ACM Intelligent User Interfaces*, 2001, pp. 33-40.

[4] Goecks, J., Shavlik, J. "Learning Users Interests by Unobtrusively Observing Their Normal Behavior," *Proc. of ACM Intelligent User Interfaces*, 2000, pp. 129-132.

[5] Grudin, J. "Groupware and Social Dynamics: Eight Challenges for Developers," *Communications of the ACM*, 35 : 92–105, 1994.

[6] Kelly, D., Belkin, N. "Display time as implicit feedback: Understanding task effects," *Proc. of ACM SIGIR 04*, 2004, pp. 377-384.

[7] Kim, J., Oard, D.W., Romanik, K. "Using implicit feedback for user modeling in internet and intranet searching," University of Maryland CLIS Technical Report, 2000.

[8] Morita, M., Shinoda, Y. "Information filtering based on user behaviour analysis and best match text retrieval," *Proceedings of ACM SIGIR'94*, Springer-Verlag, pp. 272-81.

[9] Nichols, D."Implicit Rating and Filtering," *Proc. of the 5th DELOS Workshop on Filtering and Collaborative Filtering*, Budapest, Hungary, 10-12, November 1997, pp. 31-36.

[10] Sarwar, B., Konstan, J., Borchers, A., Herlocker, J., Miller, B., Reidl, J. "Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System," *Proc. of ACM Conference on Computer Supported Collaborative Work (CSCW)*, 1998, pp. 345-354.

[11] Shipman, F., Hsieh, H., Maloor, P., Moore, J. M. "The Visual Knowledge Builder: A Second Generation Spatial Hypertext," *Proc. of ACM Hypertext*, 2001, pp. 113-122.

[12] Shipman, F., Hsieh, H., Moore, J.M., Zacchi, J.M. "Supporting Personal Collections across Digital Libraries in Spatial Hypertext," *Proc. of the ACM and IEEE Joint Conference on Digital Libraries*, 2004, pp. 358-367.

[13] Shipman, F., Price, M., Marshall, C., Golovchinsky, G., Schilit, B. "Identifying Useful Passages in Documents Based on Annotation Patterns," *Proc. of European Conference on Digital Libraries*, Springer Verlag, 2003, pp. 101-112.