EVALUATION OF A SYSTEM LAYER DESIGN FOR THE VISUAL KNOWLEDGE

BUILDER

A Thesis

by

ARUN BHARATH GOMATHINAYAGAM

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2011

Major Subject: Computer Science

Evaluation of a System Layer Design for the Visual Knowledge Builder

EVALUATION OF A SYSTEM LAYER DESIGN FOR THE VISUAL KNOWLEDGE

BUILDER.

A Thesis

by

ARUN BHARATH GOMATHINAYAGAM

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Frank M. Shipman
Committee Members,	Richard Furuta
	Ergun Akleman
Head of Department,	Duncan Walker

December 2011

Major Subject: Computer Science

ABSTRACT

Evaluation of a System Layer Design for the Visual Knowledge Builder.

(December 2011)

Arun Bharath Gomathinayagam, B. Tech, PSG College of Technology, Coimbatore,

India

Chair of Advisory Committee: Dr. Frank M. Shipman

When users are searching for documents, they must sift through a collection of

potentially relevant documents assessing, categorizing and prioritizing them based on the

current task at hand, a process we refer to as document triage. Since users' time is

precious, as much information as possible should be presented to them to aid the process

of document triage. This thesis presents a simple visualization and a set of features that

can help users in identifying information of interest. As a part of this thesis, the System

Layer of the Visual Knowledge Builder (VKB) was developed as a tab strip container.

Each of the tabs presents a different type of information about Web Documents. The

types of information currently included in VKB are: a summary of the Web Document,

keywords based on users' interests provided by the Interest Profile Manager (IPM),

popular keywords from a social bookmarking site, metadata about the Web Document, a

list of outgoing links of the Web Document, and the history of the Web Document. We

performed a heuristic evaluation to assess the usefulness of the new visualization and

features. During the evaluation, participants were asked to rate the usefulness of each of

the new web document features over a scale of 1 to 7, where a value of 1 indicated

strong disagreement, and 7 indicated strong agreement. Our results indicate that the

document summary, the keywords from IPM, popular tags, and the history of the Web

Document are expected to be most useful during the process of document triage.

DEDICATION

To my friends and parents

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Frank Shipman, and my committee members, Dr. Richard Furuta and Dr. Ergun Akleman, for their guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues, and the department faculty and staff, for making my time at Texas A&M University a great experience. I also want to extend my gratitude to all who contributed to this thesis.

Finally, thanks to my mother and father, for their encouragement and support.

NOMENCLATURE

IPM   Interest Profile Manager

tf-idf   Term Frequency Inverse Document Frequency

VKB   Visual Knowledge Builder

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE   Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Users have to interact with a large amount of information while identifying materials of interest for many tasks. For instance, when searching for content on the web, users often have to assess a large number of documents to identify those that match their needs. Document triage is the process of initially determining the value of documents (Marshall & Shipman, 1997). In the process of triage, users may use multiple applications and move back and forth between activities of searching, reading and organizing. For some tasks, this process of sifting through potentially relevant documents can take more time than completing the task at hand once the relevant documents are identified (Badi et al., 2006).

Users performing document triage can benefit from techniques that enable them to focus on the document features relevant to the task at hand. While engaging in document triage, people may indicate their interests and needs alongside the query they initially formulated to retrieve documents. Tools that interpret user interests can provide cues to identify documents of interest as the task proceeds. Prior work (Badi et al., 2006) has examined how to support triage using learned user models. Implicit and explicit indicators were used to construct user models and to provide suggestions to the users through visual cues attached to documents.

_____

This thesis follows the style of *Adult Education Quarterly*.

In addition to learned models of user interest, document triage can be supported by enhancing the query results through the display of additional document features and metadata. Where the visual cues resulting from the user models provide an indication that the system believes there is some form of relationship between resources, the effective display of document features makes it possible for the user to make such an assessment. Different tasks make different features of a document relevant to triage. For example, the user may be interested in resources which have links to either a focused or broad set of additional sites or resources which have change regularly. In such a situation, showing the contained links of the documents or the history of the document can be very useful.

This thesis describes an extension to the visualization of resources that can aid users in identifying documents of interest by presenting different features of Web Documents in the System Layer of the Visual Knowledge Builder.   The implementation has taken place within the context of the Visual Knowledge Builder 3 (VKB 3) workspace (Shipman et al., 2001). VKB 3 enables objects to have multiple layers and, in particular, VKB Web Document objects have a Main (User) Layer and a System Layer as shown in Figure 1.

Figure 1
Web document object with two object layers

The Main Layer contains a user-configurable combination of metadata (title, URL, etc.) along with a thumbnail of the Web Document. The Main Layer also enables a user to informally express preliminary interpretations of document characteristics and value via color and other visual attributes. The system uses the visual attributes of the System Layer to provide suggestions to the user without interfering with user expression (Bae et al., 2008). Overall, this visualization allows people to use perception in carrying out tasks (Vicente & Rasmussen, 1990).

This thesis develops and assesses alternative presentations for the System Layer of VKB 3's Web Document object. The new System Layer is used to show different features of a Web Document. There is no one best document feature that will help the user. The relevance of a feature depends on the user, his current task and the data. With the new System Layer, users can alter the presentation of the System Layer as they determine the relevant features of Web Documents for the task at hand.

The System Layer implemented as a part of this thesis presents various features of a document using a tab strip container. The features include relevant metadata information, keywords from the Interest Profile Manager (Bae et al., 2010), popular tags from the social bookmarking site delicious.com (Delicious, 2010), outgoing links from the Web Document, and a film strip visualization of the Web Document history. The Web Document history shows snapshots of the Web Document using the Internet Archive (Web Archive, 2010). The tab strip container provides the ability to switch between the different views while requiring relatively little screen space. It is extensible as more tabs can be added as additional valuable features of Web Documents are identified. We believe that displaying relevant features of a document using our design will aid the users during the triage process.

The following section explains the problems and issues with document triage implementations that do not account for the presentation of multiple document features. Further sections include related work, approach of this research, description, and implementation of the visualization for the System Layer. The thesis ends with an evaluation and contributions of this thesis.

CHAPTER II

PROBLEM STATEMENT

Users often interact with search engines or information visualization and organization tools to search for potential documents of interest. Such tools typically provide an overview of a collection or search results. Many of the tools today show standard document metadata such as the author of the document, size, last modified time etc. Some applications also make use of models of user interests as a basis for visualizations to guide the users to identify similar documents (Bae et al., 2008).

Apart from providing recommendations and standard metadata, additional features of Web Documents may be valuable and can also be presented. As specified earlier, different types of tasks will make different features of Web Documents relevant. Presenting different document features will allow the users to more rapidly assess documents during triage.

Search engines

Search engines are widely used to retrieve information of interest from the World-Wide Web and present a list of documents that match the user specified query terms. Users may be interested in additional information about the documents returned by the query. For example, they may be interested in the web pages that are linked to by the retrieved documents. They may also be interested in keywords that are associated with a document.

The most popular World Wide Web search engines show the documents that

match the users' queries as scrollable paginated lists. To view additional features of the documents relevant to their current task, users must navigate to each document individually. This adds to the time and effort required of information seekers. "Digging for information in the World Wide Web, as a huge, heterogeneous, unstructured load of data, represents a special challenge" (Mann, 1999). They have to scroll and navigate to different pages of search results while being involved in the process of document triage. This is often tedious and time consuming (Mukherjea & Hara, 1999). When the search results are a long list of URLs it would be good to help the user find the needle in the haystack by applying adequate visualizations (Mann, 1999). Some systems provide visualization support during the query formulation phase in the form of query expansion. The graphical query refinement visualization provided by AltaVista, initially named Visual Live Topics supports query expansion (AltaVista, 2010).

Visualization of search results often focuses on providing an overview of the set of matching documents. Trends, clusters and hotspots related to the results can be shown (e.g. clusty.com). At the document level, the overview may indicate the relevance of a single document and its relation to other documents. Such support is valuable but does not reduce the need for information seekers to navigate to individual documents to gain access to additional document features. There is a need for visualizations that provide an overview of the set of matching documents as well as access to the relevant features of individual documents. Many interfaces show titles and thumbnails for individual documents (Mann, 1999). As a part of this thesis, we address this problem by enabling users to switch between alternative document features in the visualization.

Visual knowledge builder and the system layer

The Visual Knowledge Builder (Shipman et al., 2001) uses the Interest Profile Manager (IPM) (Bae et al., 2010) to infer user interests during document triage. As mentioned earlier, the Web Document objects of the VKB have a main layer and a System Layer. In prior work, the System Layer was limited to taking the inputs from the IPM to indicate similar documents through colors of varying degrees of saturation. For example, in Figure 3, two Web Document objects which are identified by the IPM to be similar are indicated using red colored System Layers.

Though indicating similarity using colors is useful, the System Layer can do much more. In particular, it can visualize metadata and other document features which may aid the user during the process of document triage. There are a large number of features of Web Documents that may be valuable to represent in the presentation of search results. At the same time the additional information provided should not intrude the process of document triage and it should not increase the work for the user. Finally, the screen real estate for a document object's System Layer is small, meaning that the design of the individual document visualization must be spatially efficient in providing the variety of potentially valuable features to the user.

CHAPTER III

RELATED WORK

Prior research related to this thesis can be divided into the following categories: focus+context visualizations, visualization techniques for showing search engine results, techniques that reduce space contention in window based Graphical User Interfaces, and the presentation of Web Document metadata.

Focus + context visualization techniques

Mukherjea and Hara (1999) have introduced a visualization technique called Card-Vis to show search spaces to the users. The technique is based on the metaphor of a pack of playing cards and uses the *focus+context* technique to handle a large amount of information. Using focus+context techniques to show abstract information is a popular approach. Using this approach the information of interest to the user can be shown in detail while also contextual information is shown in less detail (Mukherjea & Hara, 1999). The Generalized Fisheye View (Furnas, 1986) is one such two-dimensional visualization where different regions of the 2D space are shown at different zoom factors. This technique was incorporated into VIKI (Marshall et al., 1994), the predecessor to VKB. The Perspective Wall (Mackinlay et al., 1991) visualizes linear information, integrating detailed and contextual views, by providing a bent wall which turns away from the user on the sides while providing a detailed view in the center. Document lens (Robertson & Mackinlay, 1993), table lens (Rao & Card, 1994) and

hyperbolic browser (Lamping et al., 1995) are some of the other focus + context visualization techniques.

Visualizing search results

Many systems have been developed to allow users to navigate through the result space of the World Wide Web search to find information of interest. Web search results are typically disconnected graphs and many unique interfaces have been developed to display them. SenseMaker (Baldonado & Winograd, 1997) is an interface developed for information exploration across heterogeneous sources in a digital library. Web-Book and Web Forager (Card et al., 1996) together create an information workspace for the World-Wide Web, displaying search results and embedding objects in a hierarchical 3D workspace. The Cat-a-Cone interface (Hearst & Chandu, 1997) allows for fluid two-way interaction between selections of category labels for search and display of multiple category labels within retrieval results. Websites like surfcanyon.com (SurfCanyon, 2010) provide real-time recommendations based on the user activity in addition to displaying the search results.

Mann (1999) has proposed the application of visualization techniques to four different phases of a search process namely, "(1) Formulation: expressing the search, (2) Initiation of action: launching the search, (3) Review of results: reading messages and outcomes, and (4) Refinement: formulating the next step". He discusses the usage of 'Synchronized Alternative Visualizations' such as document vector, scatterplot, bargraphs, titlebars, relevance curves and thumbnail views to facilitate the access of information space for the user.

Examples of techniques that support the initial query formulation are listing additional words from a thesaurus below the entered keywords (Veerasamy & Navathe, 1995) or displaying a network structure (Fowler et al., 1991) for queries, document content and term relations. Anick et al. describe the design of a visual interface to support Boolean query formation and modification (Anick et al., 1989). Chang (1989) discusses a similar visualization to support information retrieval from very large databases.

Space contention in window based graphical user interfaces

A major challenge to the effectiveness of window based user interfaces is the small display size of the screen. The principal obstacles to a visualization of linear information structures are (1) the large amount of information that must be displayed and (2) the difficulty of accommodating the extreme aspect ratio of the linear structure on the screen. "Window thrashing" in which the user has to spend considerable effort to keep the desired windows visible is a problem that needs to be addressed while designing applications. Many techniques have been proposed to overcome the small screen real estate problem. They fall into four categories namely, (1) alternating screen usage, (2) distorted views, (3) large virtual workspaces, and (4) multiple workspaces (Henderson & Card, 1986). Though these categories are more relevant for representing different applications in the small display screen the concepts can be applied in displaying multiple objects like the user and System Layers in the VKB.

The Xerox Star (Smith et al., 1982, April) provides file drawers where documents can be stored, fetched and reopened as needed. It makes use of alternating

screen usages for the purpose. Distorting objects in the workspace is another way to gain space. The usage of icons is one of the oldest techniques for distorting views (Pygmalion, 1975). Smalltalk was an early system to use icons to display objects and many other systems followed. Windows are typically shrunk into small pictures that would remind the user of the original window. Overlapping windows (Goldberg, 1984) is also another technique to distort objects. Smalltalk uses this technique as well. Windows are allowed to cover each other leaving only a portion to remind the user of the original window. The Boxer (diSessa, 1985) system is a different distorted view system which makes use of a spatial box metaphor to represent many semantic notions of containment.

Systems such as Sketchpad (Sutherland, 1963), Dataland (Bolt, 1984) and Cedar Whiteboard system (Donahue & Widom, 1986, January) use large virtual workspaces and they provide zoom-in and zoom-out capabilities to display the objects.

Rooms (Henderson & Card, 1986) is a window manager that overcomes the small screen real estate by exploiting the statistics of window access, dividing the user's workspace into a group of virtual workspaces and providing transitions between them.

The existing Web Document object implementation in VKB 3 uses overlapping objects to show both the user and System Layers. The System Layer implementation proposed as a part of this thesis will keep the current behavior of overlapping objects to show both the user layer and part of the System Layer or vice versa based on which layer is selected to be on top. The concept used is similar to the distorted views technique.

Web document metadata

Closer to the focus of this thesis is the display of Web Document metadata. VKB already displays limited metadata in the main view of the document object but this is limited to the title and URL. Systems such as DocuBrowse (Girgensohn et al., 2010) use additional metadata to enable faceted browsing and searching of document collections. Gheel and Anderson have analyzed the usefulness of activity metadata for enhancing Web Document retrieval. Their results indicate that Web Document activity metadata helps the users to remember and retrieve previously viewed web page documents (Gheel & Anderson, 2007). They also discuss the benefits of using document metadata that exists as part of the document and those that are also generated by its use, to assist users in remembering the document content and retrieving it. They claim that "by extracting valuable descriptive information from a document, and analyzing what a user does, such as the files they open, when they are used and for how long, a repository of information can be built up to help users with the tasks of finding and reminding". They also emphasize the need for a visualization to elicit meaning from the vast raw metadata that can be acquired (Gheel, 1999).

The contribution of this thesis is the implementation and evaluation of a visualization for a wide variety of Web Document features. These features go beyond traditional metadata. The implementation allows users to select among these features and displays them in the System Layer. The additional Web Document features can help users to find useful documents and to refine their understanding of their task and search queries. Users can submit queries, see the results, judge the relevance of the Web

Documents using the document features and adjust the search queries to improve the

results.

CHAPTER IV

PRIOR WORK

The Visual Knowledge Builder (VKB) is a spatial hypertext system that enables users to retrieve and visually arrange potential documents of interest. Much of the prior work on supporting triage has focused on visualizations based on inferred user interests.

VKB uses the Interest Profile Manager (IPM) to infer user interests during document triage. VKB 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. When 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. The communication between VKB and IPM is two way. VKB sends information to the IPM as input to the algorithm that computes user interests. The IPM infers user interest based on information collected across all of the triage applications and sends the inferred interest back to VKB and other registered applications. The IPM will retrieve the documents (using their URLs), aggregate their term vectors, and identify the metadata that is common to all similar documents as indicated by the user. This processing results in a characterization of the document class.

IPM then calculates an interest rating for each document. It does so by combining the user activity associated with each document across triage applications. Finally, all the individual interest values of each of the documents are aggregated to

quantify the user's interest in the document class. "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)". Figure 2, illustrates the interaction between the IPM and other triage applications.

Figure 2
Communication between triage applications
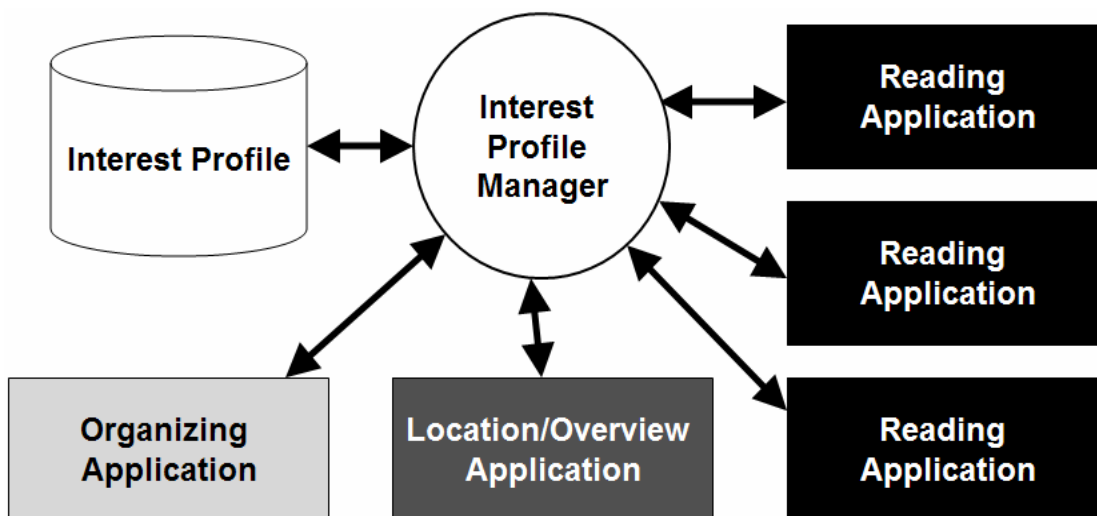
If a document is close to one of the classes, the System Layer's color changes so that it matches the document objects in that class to indicate to the user the similarity between the document and others the user has examined. The color's saturation reflects the system's confidence (the degree of match between the document's term vector and the class term vector) (Bae et al., 2010). Figure 3, shows two similar documents as

indicated by the red colors of the respective System Layers. The colors vary in saturation

as they reflect the system's confidence in the similarity measure.

Figure 3
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 other triage applications
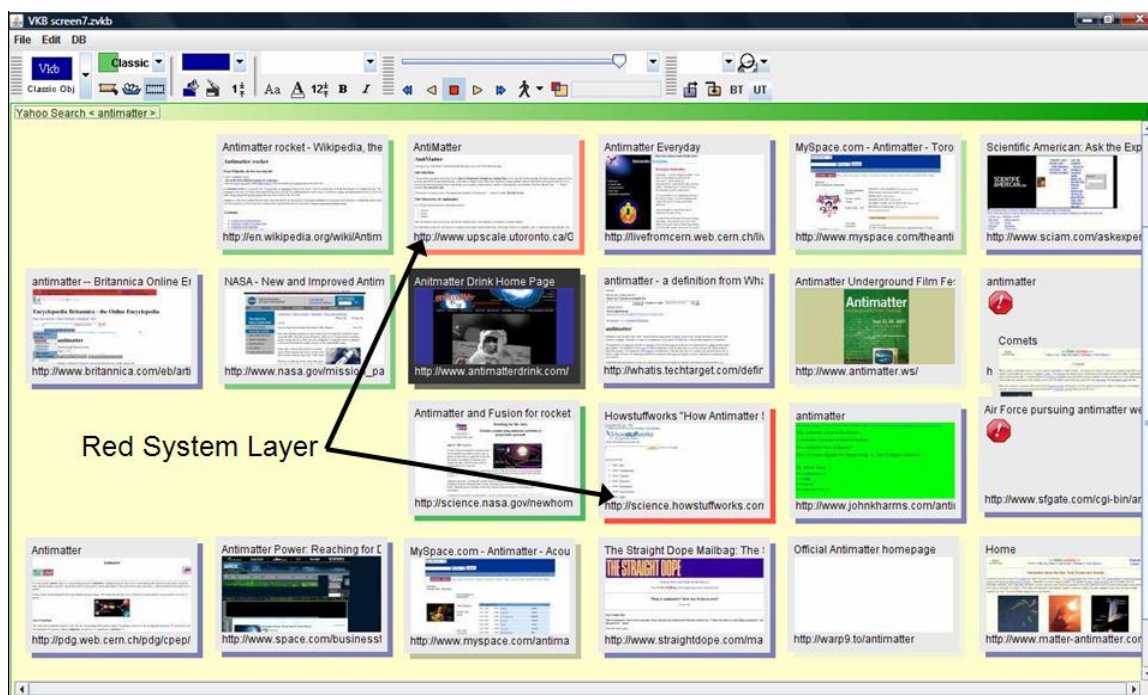
CHAPTER V

APPROACH AND DESIGN

The previous sections introduced the current state of the System Layer implementation and the visual cues it provides. As mentioned in the prior work, the System Layer provides color cues and the summary of each of the Web Documents. The summary of a Web Document is shown when the user switches to the System Layer from the user layer. This use of the System Layer can be greatly expanded and enables the display of most any type of information about Web Documents, including most types of traditional metadata and/or other features. The following section explains the approach for the implementation and evaluation of the new System Layer visualization in VKB 3.

Visualization

Since the System Layer occupies a small screen space in the VKB workspace, showing a lot of information at once will render the object cluttered and the user will have a hard time finding the relevant information. The visualization should be simple and it should enable the user to pick and choose the desired type of information with relative ease. Additionally, its interaction should not add undue complexity.

Considering these factors, the System Layer is implemented as a Tab Strip container. The tab strip provides the advantage of listing all the view headers at once. Thus users have a stable context and can navigate easily between the views. Tab strips are also the ideal choice for presenting multiple views of information when the views

look very different from one another. Users can switch between views rapidly allowing

the visualization to be adapted to their changing needs.

Prior to the work on this thesis, VKB 3 allowed the user to switch between the

System Layer and the User Layer using the "Switch Layer" option from the context

menu as shown below in Figure 4. As part of this thesis, a new Web Document Toolbar

was created (explained below) that allows users to toggle between the user and System

Layers and to switch the System Layer's view between various features of the Web

Document.

Figure 4
"Switch layer" option in the context menu of the web document object can be used
to switch between user and system layers
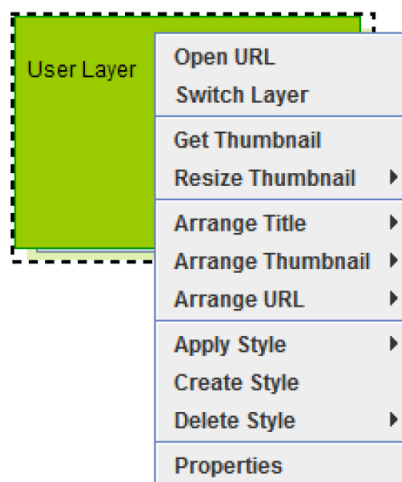
System layer as a tab strip container

Once the user switches to the "System Layer", it shows a tab strip container without tab headers by default. Since the size of the Web Document object is small, navigation between different views using tab headers will be difficult. Instead a drop down menu with the view titles is provided. Users can use the drop down menu to navigate between the views with relative ease. Figure 5, shows the visualization of the System Layer as a tab strip container with a drop down menu. The menu items indicate the type of information that will be displayed in the corresponding view.

Figure 5
The system layer as a tab strip container with a drop down menu for supporting navigation
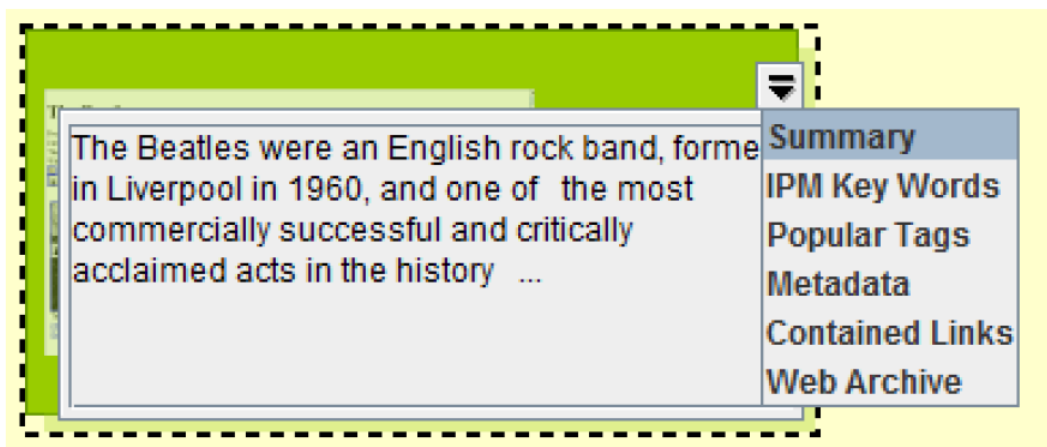
Users can resize the Web Document object. When the size of the object is large enough to display tab headers, they will be shown for all the tabs. Figure 6 shows a

resized Web Document object with enough space to display all the tab headers. The drop

down menu will also be available for navigation. Tooltips are enabled for all the views.

They display the whole content of a view. This enables the user to see all of the content

of a view without scrolling. Scrollbars are automatically enabled for views whose

content exceed the view size.

Figure 6
System layer with tab headers enabled for navigation

Web document toolbar

Since the VKB workspace can have many Web Document objects, users should

be able to toggle between the User and System Layers for multiple objects as once. If

the interaction involves a lot of effort then it will affect the process of document triage.

They should also be able to resize the Web Document objects easily to visualize and

consume the presented information.

Figure 7 shows the Web Document toolbar as a part of the VKB toolbar. Figure 8 shows the popped out Web Document toolbar. The toolbar allows users to toggle between System and User layers, to choose a view of the System Layer and to resize the Web Document object with few clicks. These capabilities reduce the effort of users to make use of the various document features available via the System Layer.

Figure 7
Web document toolbar
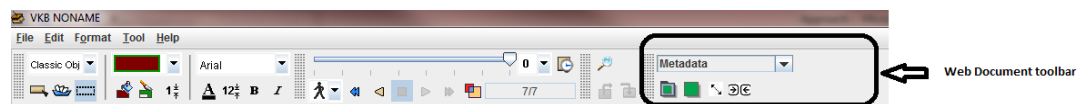
Figure 8
Web document toolbar (popped out)
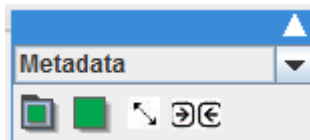
A description of each of the controls of the toolbar is given below:

Switch to the System Layer - Clicking on this icon will switch the current layer of the Web Document object to the System Layer.

Switch to the User Layer - Clicking on this icon will switch the current layer of the Web Document object to the user layer.

Enlarge the web document object - Clicking on this icon will resize the Web Document

object such that it can display all the tabs of the System Layer.

Reduce the web document object to the default size - Clicking on this icon will resize the

Web Document object to its default size.

Web document features

The new System Layer implementation provides a list of Web Document

features. Most of the features are either extracted from the Web Document itself or they

are identified using various sources from the web. The features are shown using the tab

strip visualization described earlier.

The implementation and description of the System Layer features are presented

in the next chapter.

CHAPTER VI

SYSTEM LAYER VIEWS DESCRIPTION

The new design of the VKB document object System Layer provides for viewing a

number of document features. This chapter describes each of these features, including

how their content is generated and how it is displayed in the System Layer. The System

Layer provides access to the following types of information for a Web Document object.

- **Summary** – Short summary of the content of the Web Document.

- **Interest Profile Manager keywords** - Keywords provided by the Interest Profile

  Manager.

- **Popular tags** - Popular tags from the social bookmarking website,

  'delicious.com'

- **Metadata** – Information such as page rank, page size, and the number of

  outgoing links.

- **Outgoing links** – List of all the outgoing links.

- **Web Archive** - History of the Web Document using the internet archive.

These features are described in the following sections.

Summary

As specified in prior work, prior to this thesis the System Layer showed only the

summary of Web Document objects. The summary is now shown as a view in the new

tab strip container (see Figure 9) implementation of the System Layer. The summary is

meant to provide an overview of the content of the Web Document object. It can be

useful when users are selecting and categorizing search results from the web. Users can validate the search results by identifying those documents concerned with alternate meanings of the search terms and refine the search queries if required.

Figure 9
The system layer displaying the 'summary' of a web document object
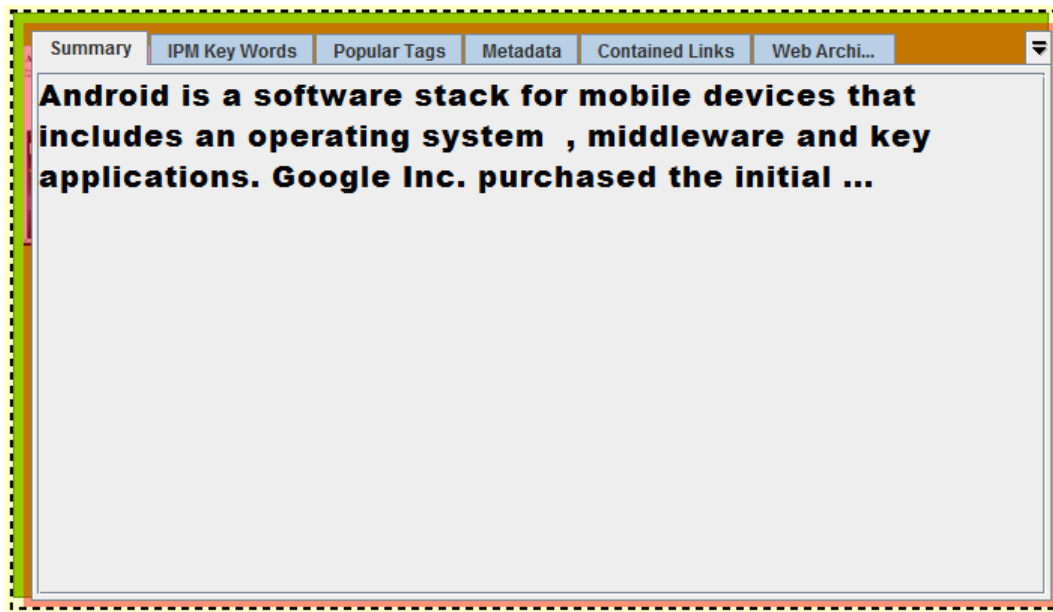
Interest profile manager keywords

The IPM maintains user interest profiles and it generates a set of keywords for every Web Document in a user's VKB workspace. These keywords are generated considering the user's interest and past activity. Thus, these terms are expected to be relevant to users and their task. Presenting the IPM keywords found in each document provides a personalized view of the vocabulary of the documents. This view is expected to help the user to sift through documents to locate those most relevant to their task. For

example, consider a scenario where the user's workspace contains documents related to popular music bands and the user is specifically interested in British music bands due to his current task, the IPM may generate keywords such as British, Beatles, Eagles, etc., for documents. Because the keywords are based on the user's more general interest in music, the keywords will help to identify which documents are about British music and which are about music in general or music from other countries.

Figure 10
The system layer displaying keywords from a web document using tf-idf, which also considers documents from the user's interest profile
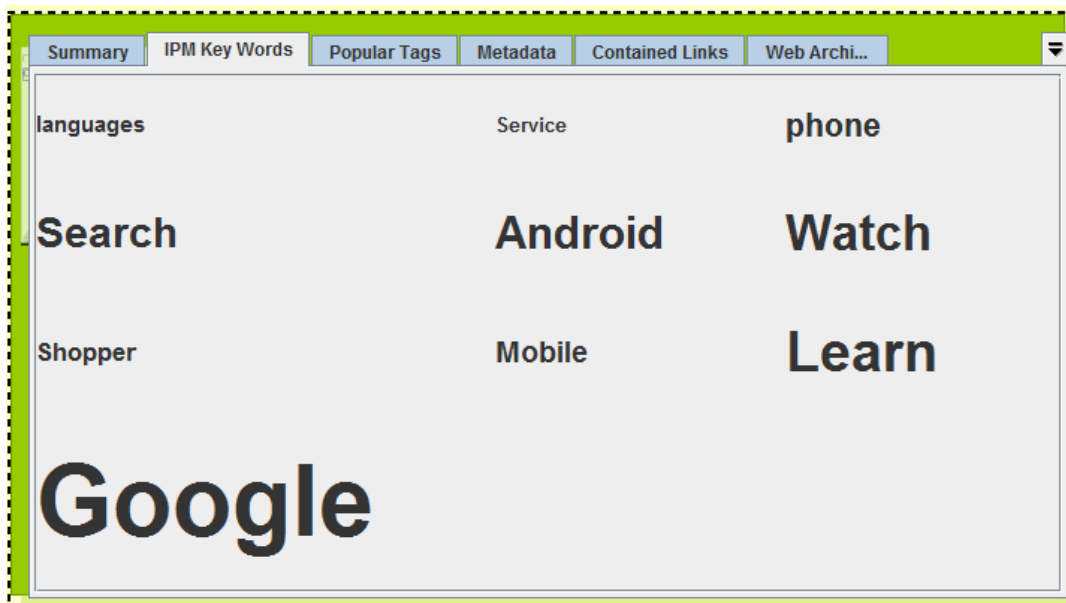
The Interest Profile Manager uses the term frequency-inverse document frequency (tf-idf, 2010) weight to specify how important a word is to a document in a collection. It considers all the documents from the user's interest profile while

calculating tf-idf weights. Words which are unique to a Web Document and which have higher weights appear with a bigger font size to indicate their significance. This visualization is very similar to the "word cloud" visualization provided by wordle (Wordle, 2011). In Figure 10, the keywords "Google", "Android", "Watch" and "Learn" are more prominent than the other keywords. Using this information, the user may interpret that the website provides demonstrations of Google Android mobile phones.

The users can also create a new collection of Web Documents using all or some of the IPM generated keywords as a search query.

Metadata

Information from META tags of a Web Document such as author, content etc. are shown in the System Layer in this view. Also the page rank, size and the number of outgoing links of a Web Document are shown. The size of a document and the number of outgoing links may give an idea about the length and the type of the content.

Figure 11
The system layer showing the 'metadata' of a web document object
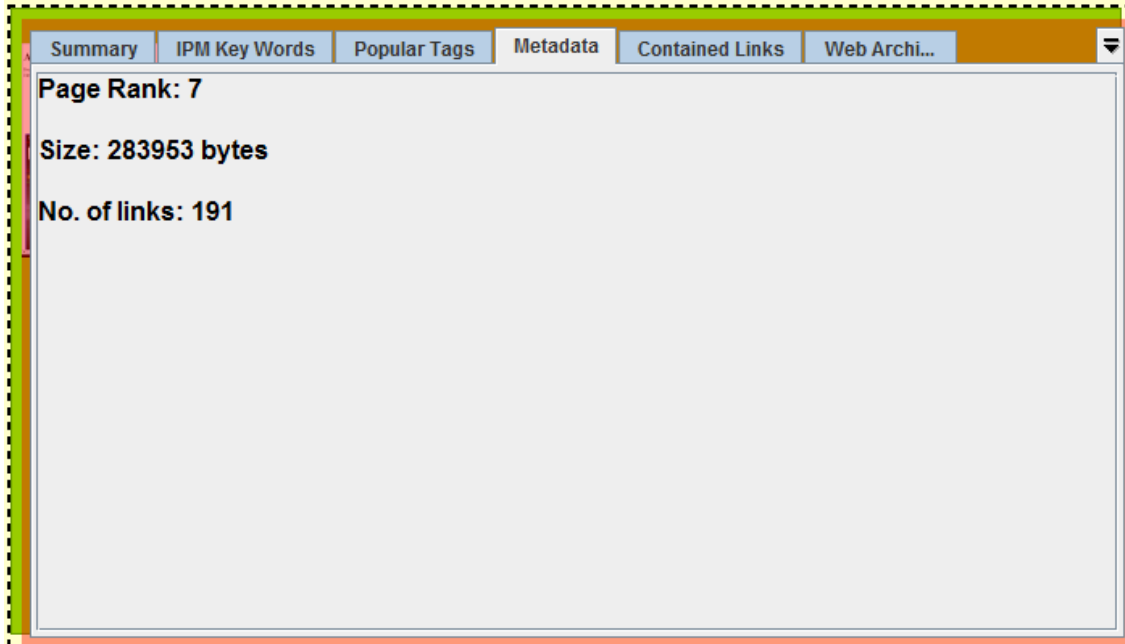
Page rank of a Web Document is calculated using Google Toolbar API. Page

rank is a whole number that ranges from 0 to 10. The most popular websites have a

Page Rank of 10 (PageRank, 2010). Based on the Page Rank of a Web Document, the

user can sense its popularity. This will help the user to validate the search results. If

most of the Web Documents in a search result have an unexpectedly low page rank value

then the user may decide to refine the search query terms. A snapshot of the System

Layer displaying the metadata of a Web Document is shown above in Figure 11.

Popular tags

Delicious (Delicious, 2010) is a social bookmarking website in which users can tag each of their bookmarks with freely chosen index terms. The most popular tags for a Web Document, if available, will be shown in the System Layer in this view. These tags, like the IPM generated keywords can prove to be significant indicators of documents of interest to the user. While the tags are not personalized to the individual user, the tags have been selected by people who looked at the document rather than selected by an algorithm.

For example, tags such as 'android', 'google', 'apps', 'mobile' and 'phone' as shown in Figure 12, provides a good indication about the Web Document. From these tags, the user can immediately identify that the website provides Google mobile applications for android phones. Using this information the user can identify documents of interest from a collection.

Figure 12
The system layer displaying popular tags from the social bookmarking site,
delicious.com, for a specific web document object
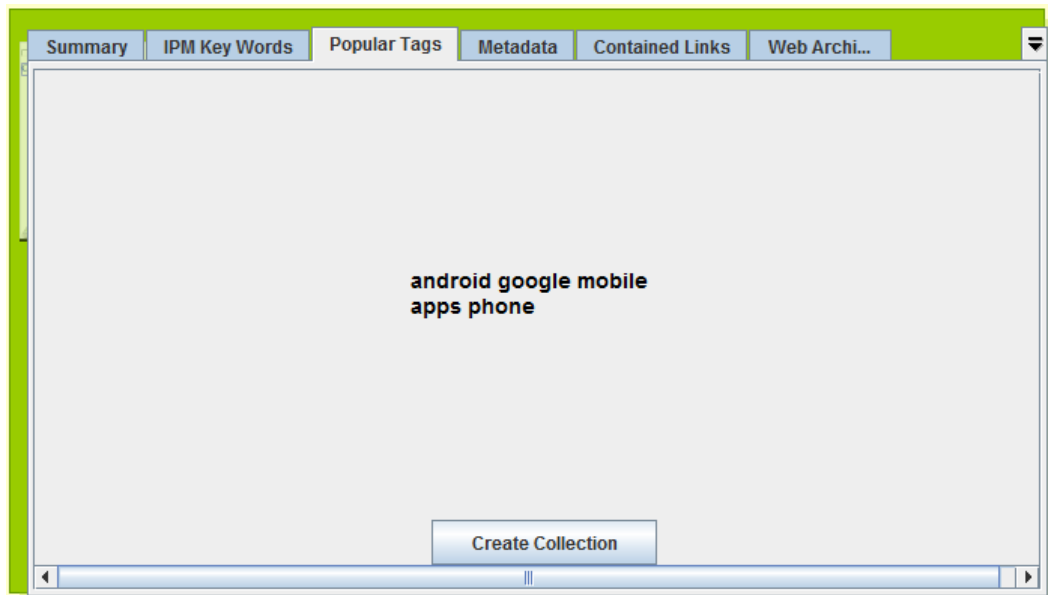
A snapshot of the System Layer showing the most popular tags of a Web

Document is shown above in Figure 12. A "Create Collection" button is provided which

allows the users to create a new collection of Web Documents using the popular tags.

The popular tags will be formulated into a search query and the resulting Web

Documents will be created inside a collection when "Create Collection" is used.

Contained links

The outgoing hyperlinks contained in a Web Document can be of interest to the

user. The System Layer will show the contained links of a document as shown in Figure

13. If there are a lot of links then the user can scroll or use the tooltip to see them.

Listing all the outgoing links of a Web Document may indicate the type of Web

Document. For example, documents with links to lots of sites, such as blog aggregator sites, will easily be distinguished from documents with no links and documents with links to a single site.

Figure 13
The system layer displaying the contained links of a web document object
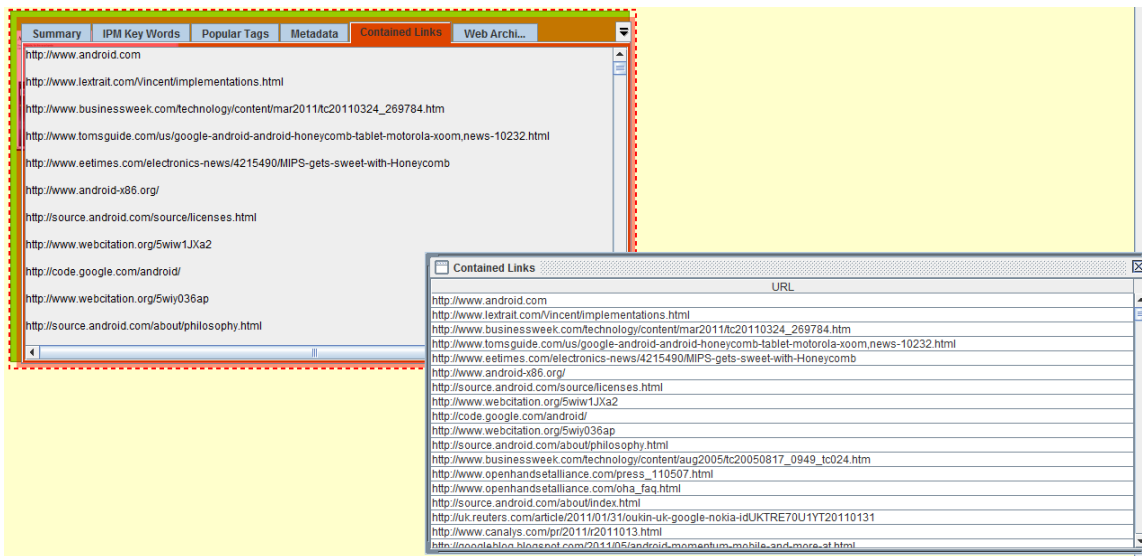
Web document history

Using the Internet Archive (Web Archive, 2010), the history of a Web Document can be tracked. This view of the System Layer shows thumbnails of archived instances of a Web Document. With the thumbnails, the user can see the evolution of a Web Document and its history. A snapshot of a Web Document history as displayed in the System Layer is shown below in Figure 14.

Figure 14
System layer displaying thumbnails of a web document object's history from the web archive
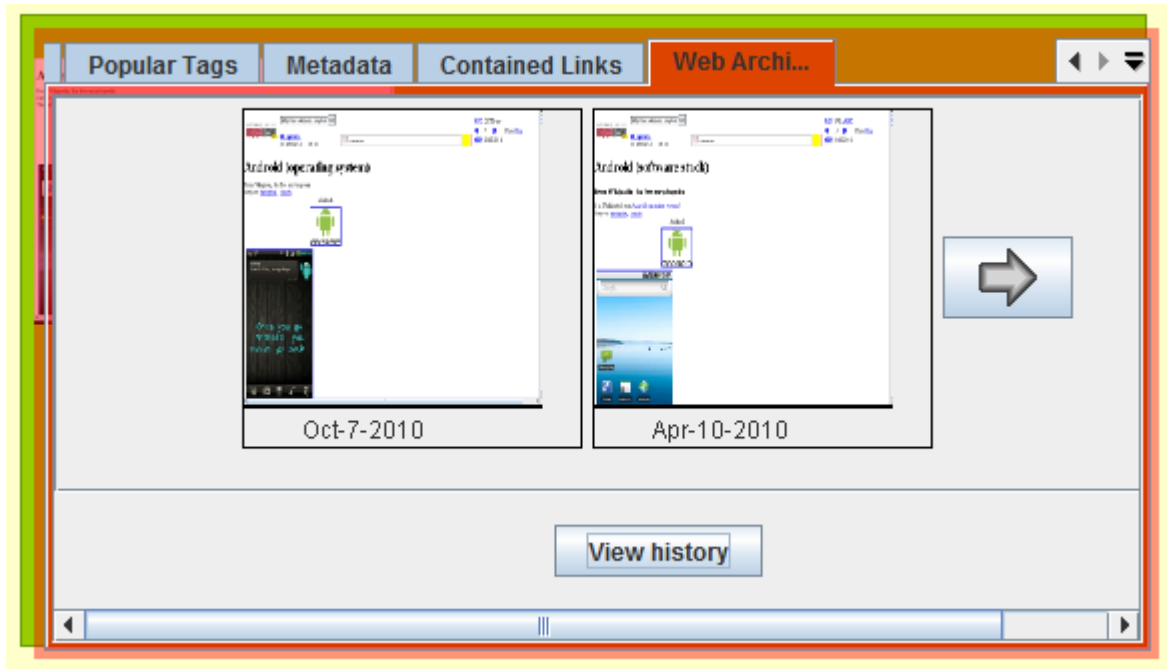
One difficulty for this view is that the Internet Archive (Web Archive, 2010) has a large number of snapshots for many websites. When displaying the snapshots of Web Documents from the history, showing all of them may not be useful. Users will be inundated with information. In order to choose candidates from a corpus of snapshots, automatic history clustering, as described by Kim and Shipman (2010), is used. Snapshots that are closer to each other based on time are grouped as clusters and a single snapshot from each of the clusters is chosen and displayed in the System Layer. The snapshot that was archived the latest is chosen as the representative from a cluster. The

clustering makes sure that the chosen snapshots are significantly distanced apart with respect to time. This selection process reduces the risk that the user is overloaded with snapshots and aims to select a set of snapshots that will prove to be valuable.

The ability to compare the contents of two snapshots of a Web Document from the Internet Archive is provided within this view. Users are able to select two snapshots of a Web Document and compare their content using a context menu as shown in Figure 15. The result of the comparison is shown as an HTML document with content additions and deletions marked appropriately as shown in Figure 16. The added contents are highlighted in green and the removed contents are highlighted in pink with the text displayed with a strikethrough.

Figure 15
System layer with two snapshots of a web document being selected for content comparison
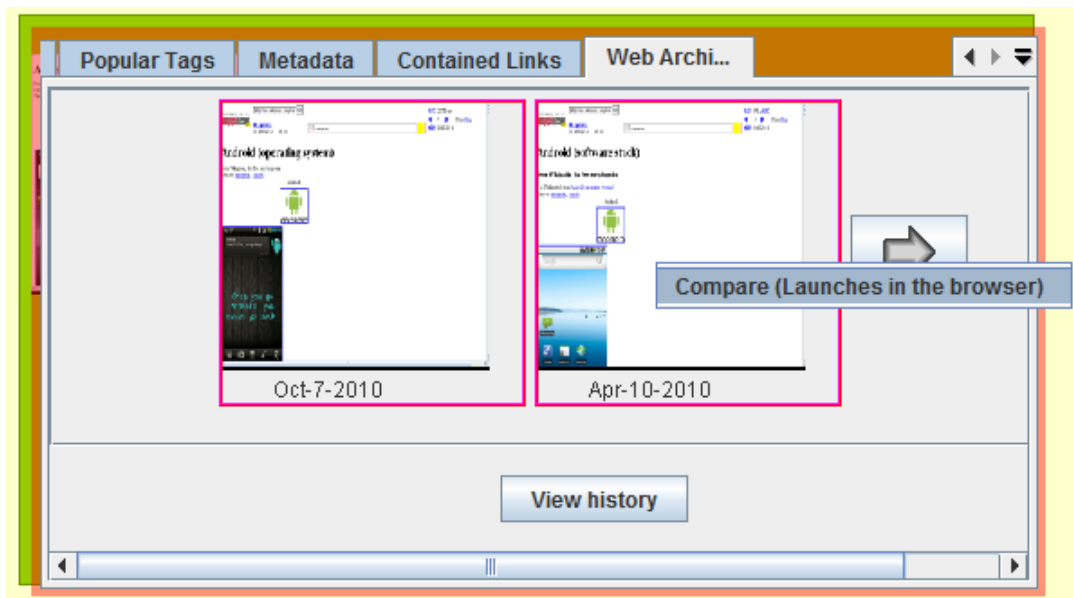
Figure 16
Result of a web document snapshots' comparison shown as a html
document

79. ^
"When will other Google products support WebM and VP8?". *The WebM Project*.
http://www.webmproject.org/about/faq/#webm_video_file_format. Retrieved 2010-05-20.

80. *a b* "Android 2.3 Platform Highlights". *Android Developers*. 6 December 2010.
http://developer.android.com/sdk/android-2.3-highlights.html. Retrieved 2010-12-07.

Document history can give an idea about the changes that a Web Document has gone through over time. For example, a Wikipedia article is likely to be edited and changed frequently. Using the filmstrip visualization and the compare snapshots ability provided in the Web Archive view, the user can get an idea about the changes occurred and when.

Consider a scenario where a user creates a VKB workspace with several Web Documents, saves it and revisits it after several months. Using the Internet Archive, the user can see how a Web Document has evolved since the previous access. Using this information, the user can reassess the usefulness of those Web Documents that have changed since their inclusion in the space.

Summary

The new VKB Web Document object System Layer provides access to a number of document features that may be of use during document triage but were previously unavailable. The next chapter describes a qualitative evaluation of the interface to and content of the views to assess the potential of the new capabilities to aid real users.

CHAPTER VII

EVALUATION

The process

A heuristic evaluation of the System Layer interface and its features was performed to assess the potential value of the new information provided within the VKB System Layer. As mentioned earlier, the features include a document summary, IPM (personalized) keywords, popular tags from delicious.com, metadata, contained (outgoing) links, and the history of the Web Document.

A total of 10 participants evaluated the user interface and the Web Document features that are shown in the System Layer. The participants were computer science graduate and undergraduate students. They were initially given a demo of the System Layer user interface and each of the features. After the demonstration, the participants were the given the option of using the Visual Knowledge Builder with the new System Layer implementation. The participants were then asked to fill up a questionnaire (see Appendix 1). They were asked to rate the usefulness, effectiveness, ease of use, ease of learning and their overall satisfaction of the System Layer design and its features. Their responses were recorded using a Likert scale where a value of 1 indicated strong disagreement, and 7 indicated strong agreement. Participants were also asked to provide their input on when a feature will be useful and suggestions on how to improve each of the features. They were also asked to list the most positive and negative aspects of the design and features. All the suggestions and comments are provided in Appendix 2.

The following discussion explores the results of the evaluation in terms of the usefulness of each of the features, the usability and overall satisfaction of the System Layer implementation. The participants took approximately 20 to 25 minutes to perform the evaluation.

Results and analysis

The results and analysis of the evaluation are presented for each of the features below. Participants' ratings of the ease of use, ease of learning of the user interface and features, and their overall satisfaction are also discussed.

Document summary

As shown in Table 1, the results indicate that the participants found the summary of Web Document objects to be useful for document triage. They felt that summary helps them to get a quick look at the content of the Web Document and it helps them to quickly find if a Web Document is interesting to them.

Table 1
**Mean and standard deviation of the responses about the usefulness of 'summary'
based on a Likert scale (1 to 7)**

	Summary	
	Average	Standard Deviation
1. It is useful	6.3	0.823272602
2.It helps me be more effective	6.5	0.527046277
3.It helps me be more productive:	6.1	0.567646212
4.It saves me time when I use it:	6.2	0.918936583
5.It meets my needs.	5.8	1.032795559

One of the participants mentioned that summary would be useful to validate results of search queries. By looking at the summary of the Web Documents, users can quickly identify documents of interest. Some felt that summary would be useful to categorize useful documents.

One of the participants felt that the summary is too small. Three of them preferred a more intelligent summarizer. One of the suggestions was to show some of the important terms of a document using term frequency apart from providing a text snippet. Overall participants found the summary to be useful for document triage.

Interest profile manager keywords

As expected, the participants found the keywords generated by the Interest Profile Manager to be useful. Since the keywords are generated considering users interest profiles, they were found to be significant indicators of usefulness of documents. The average and standard deviation of the participants' responses are shown in Table 2.

Table 2
Mean and standard deviation of the responses about the usefulness of 'ipm keywords' based on a Likert scale (1 to 7)

	IPM keywords	
	Average	Standard Deviation
1. It is useful	6	0.942809042
2.It helps me be more effective	6.2	0.918936583
3.It helps me be more productive:	6	0.942809042
4.It saves me time when I use it:	5.9	1.197219
5.It meets my needs.	5.6	1.505545305

One of the participants mentioned that it is useful as "it gives you a pictorial assessment of the frequency of words in the content and helps in estimating the relevance of a Web Document to your search". One of the responses for when IPM keywords will be useful was:

"A frequent user of the VKB will find it more useful when the system suggests the keywords/topics of discussion in the document. These personalized suggestions helps in helping the user in finding out the topics of discussions in the document looked from his/her perspective. It helps the user in finding out new knowledge (keywords) that help in refining his search terms/criteria. It also helps the user in knowing what his interest trends have been in history looking at the interest terms and their magnitude."

The above response clearly summarizes the need and usefulness of system generated keywords when one is performing document triage.

Some of the participants felt that some of the keywords were not strongly relevant to the search. This suggests that there is a potential to improve the algorithm that selects the keywords to present.   One of the participants felt that adding animation or color coding while displaying the keywords will be useful. Most of the participants liked this feature and they felt that they need it.

Popular tags

Participants found the popular tags to be equally useful to the IPM keywords, if not slightly more useful. Since the popular tags are from delicious.com, they were found to be very good indicators of the type of a Web Document. These tags can help users to

categorize a document very easily. The results of the participants' responses are shown

in Table 3.

Table 3
Mean and standard deviation of the responses about the usefulness of 'popular tags' based on a Likert scale (1 to 7)

	Popular tags	
	Average	Standard Deviation
1. It is useful	6.2	0.788810638
2.It helps me be more effective	6.3	0.948683298
3.It helps me be more productive:	6.3	0.823272602
4.It saves me time when I use it:	5.9	1.100504935
5.It meets my needs.	5.9	0.737864787

One of the participants mentioned that this view was very useful as it shows tags

that other people have used for a Web Document. Since most of the delicious.com

community uses these tags for a web site, it is a significant indicator of the type of

content of a web site. Another participant mentioned that these tags can be considered

as suggestions from other users. The below response from a participant summarizes the

need for popular tags.

"It helps the user in finding out what the rest of the world is thinking

about the document and the favorite topics of interest related to the document.

This also helps the user in getting a broader understanding of the document and

generally helps in refining his/her search criteria resulting in more quickly

finding the required information."

In response to suggestions for improving the popular tags, participants said that

showing how strong each of the tags is for a Web Document will be good. They also

suggested using more online resources to compute the tags for a Web Document.

Metadata

Overall, the metadata view was considered the least useful. That said,

participants found the page rank to be useful. They found the page rank to be indicative

of the confidence level of search results. One of the participants mentioned that, "when

viewing a higher ranked page, it gives you a satisfaction of going through a more

relevant page to your search". Another participant felt that metadata such as creation

date and modification date is useful in assessing the age of a Web Document.

Participants' average rating about the usefulness of metadata is shown in Table 4.

Table 4
Mean and standard deviation of the responses about the usefulness of 'metadata'
based on a Likert scale (1 to 7)

	Metadata	
	Average	**Standard Deviation**
1. It is useful	5.2	1.316561177
2.It helps me be more effective	5.3	1.33749351
3.It helps me be more productive:	5.1	1.523883927
4.It saves me time when I use it:	5.1	1.286683938
5.It meets my needs.	4.9	1.663329993

Some of the participants had doubts whether a novice user would understand what page rank actually means. The size of the document helps in estimating the length of the document to read. Some of the participants were not were sure about how the size of a document and the number of outgoing links would be useful. One of the suggestions was to standardize the metadata which poses a challenge as currently the data is whatever is attached by the document author plus the computed metadata. Standardizing the display would likely remove the document author-provided content.

Contained links

Contained (outgoing) links were found to be useful when the users want to visit a related link directly without examining the Web Document. Participants' average rating about the usefulness of 'Contained links' is shown in Table 5. One of the participants specified that showing all the outgoing links makes it easier to reference related links.

Table 5
Mean and standard deviation of the responses about the usefulness of 'contained links' based on a Likert scale (1 to 7)

	Contained links	
	Average	Standard Deviation
1. It is useful	5.6	1.505545305
2.It helps me be more effective	5.7	1.251665557
3.It helps me be more productive:	5.6	1.505545305
4.It saves me time when I use it:	5.8	1.475729575
5.It meets my needs.	5.5	1.178511302

Some of the participants did not find the contained links to be very useful. Some of the suggestions were to show the title of the outgoing links instead of displaying the URL, showing only the top 'n' outgoing links based on page rank, allowing the user to quickly create a VKB collection out of the outgoing links to get more information about the links and showing more metadata with each link such as the title and summary terms.

Web archive

As the results shown in Table 6 indicate, participants found the 'Web Archive' to be the most useful feature. The Web Archive helps the user to see the history and evolution of a Web Document. Participants found the 'compare snapshots' feature to be very useful. Users can find the version history and compare the contents between different versions using the Web Archive. This is also helpful when the content of the pages are changing frequently.

Table 6
Mean and standard deviation of the responses about the usefulness of 'web archive' based on a Likert scale (1 to 7)

	Web Archive	
	Average	Std Dev
1. It is useful	6.5	0.527046277
2.It helps me be more effective	6.2	1.229272594
3.It helps me be more productive:	6.1	1.286683938
4.It saves me time when I use it:	6.4	0.966091783
5.It meets my needs.	6.5	0.707106781

Some of the responses include,

> "Historical information can indicate activity and validity of information.
> Useful for news or modification of source code etc."

> "This sure helps in understanding how the current document or the content inside
> the document has evolved across time. This is more useful while researching on
> frequently changing topics and could help in easily pointing the changes and
> analyzing their causes quickly."

When responding to suggestions to improve this feature, participants felt that it would be nice to also archive snapshots of the users' view of the pages, showing a brief summary of the updates through thumbnails and providing the ability to compare archive from a particular day or period.

Ease of use

Participants found the user interface and the features easy to use. Most of them found the tab strip to be simple and convenient to use as indicated by the average ratings shown in Table 7. One of the participants commented that, "it is quite easy to use and most of the information presented is easily understandable without needing documentation". The evaluators found navigating between tabs easier with the Web Document toolbar as it allows switching tabs of multiple Web Document objects with a single selection. One of the suggestions was to provide more customization by allowing the users to disable features and enable only the desired ones.

Table 7
Participants' rating of the 'ease of use' of the system layer and its features based on a Likert scale (1 to 7)

	Average	Std Dev
1. It is easy to use.	6.4	0.699205899
2.It is simple to use.	6.3	0.823272602
3.It is flexible.	5.8	1.032795559
4.Using it is effortless.	6	1.154700538
5. I can use it without written instructions.	5.4	1.712697677
6. It is user friendly.	5.9	1.286683938

Ease of learning

The evaluators felt that learning to use the interface and features was straight forward and easy. All of them found it easy to remember how to use it. Table 8 shows the average ratings for the ease of learning of the new interface and features.

Table 8
Participants' rating of the 'ease of learning' of the system layer and its features based on a Likert scale (1 to 7)

	Average	Standard Deviation
1. I easily remember how to use it	6.5	0.707106781
2.It is easy to learn to use it.	6.8	0.421637021

Satisfaction

The majority of the participants were satisfied with the new System Layer implementation and its features. They felt that the new features will help in the process of document triage. The average of their ratings on their overall satisfaction of the interface and its features is shown in Table 9.

Table 9
Participants' rating of the 'satisfaction' of the system layer and its features based on a Likert scale (1 to 7)

	Average	Standard Deviation
1. I am satisfied with it.	6.2	0.788810638
2.I would recommend it.	6.5	0.527046277
3.I feel I need to have it.	6.3	1.33749351

Positive and negative aspects

Participants were asked to state the most positive and negative aspects of the new System Layer and its features. Some of the responses are quoted below. A complete list of the responses can be found in Appendix 2.

"I liked the Web Archive function and summary the possibility to see different features in different documents."

"Faster search and personalized search. It can surely help to decide faster on the search results."

"I liked the one common advantage in all the features of how they help the user in finding a lot of information about the document even before opening it. It saves lot of time."

"Some features may be more useful with minor modifications - Ex: Title of the page instead of links."

"1. Missing a few visual cues

2. Metadata may not be very effective in most cases

3. Concepts like online social bookmarking may overlap with these features, throwing the usefulness of this tool into a risk."

"1. I find many more usefulness of entire tool and also the System Layer in specific

2. Easy navigation between tabs

3. Adds more value for my interests. Makes me understand my interests better

4. Opens up ways of building my interests on unexplored related areas."

Overall, the evaluation indicates the interface and content of the System Layer were perceived as valuable to participants. Of the features, the document history, document tags, personalized keywords, and summary were viewed as more likely to help while the metadata and links out features were viewed as less likely to be of value.

CHAPTER VIII

CONCLUSION AND FUTURE WORK

Finding desired information on the web is a problem that grows by the day as the web keeps getting bigger. Users do not have an unlimited amount of time and the time users have is too precious to be spent in brute force searching. This thesis presents a simple visualization and a set of features that can aid users in finding information of interest. The System Layer of the Visual Knowledge Builder was developed as a tab strip container that displays different types of information associated with a Web Document. The features are intended to support the user in quickly identifying and categorizing Web Documents. The features include a summary of the Web Document, keywords from the Interest Profile Manager, popular tags from delicious.com, metadata associated with a Web Document, list of outgoing links of a Web Document, and the Web Archive feature which can be used to track the history of a Web Document. A Web Document toolbar was also added to the VKB to facilitate interaction with Web Document objects.

A heuristic evaluation of the new System Layer user interface and its features was performed. The participants found the document summary, IPM keywords, popular tags and the Web Archive features to be the most useful. They also found the visualization and the user interaction to be simple, easy to use and easy to learn. Web Documents have a lot of information associated with them that can be used to aid document triage. We showed only a subset of them and evaluated them. The evaluation

results have indicated that there is room for improvement. Participants suggested extensions and improvement to the existing features and as well new features worth adding.

Currently the IPM keywords are generated using tf-idf which also considers documents from users' interest profiles. This algorithm can be extended or modified to incorporate more parameters from users' interests. As participants suggested, simply showing the list of outgoing links is not very useful. Possibly showing a subset of the outgoing links with additional information about them can be useful. The summary shown is currently is too small for many uses. In the future, the summary could be computed using an intelligent summarizer, which can provide more information about the type of content of a Web Document. In addition to the features that were discussed, it was pointed out that more useful information about Web Documents can be computed using other content on the web.

Document triage is a time-consuming activity when gathering content. The new document object System Layer points towards the potential of providing easy access to a wide range of document features to aid in this activity.

REFERENCES

AltaVista. (2010). Visual live topics. Retrieved 2010, from http://www.altavista.com.

Anick P. G., Brennan J. D., Flynn R. A., Hanssen D. R., Alvey B., & Robbins J. M. (1989). A direct manipulation interface for boolean information retrieval via natural language query. In *Proceedings of the 13th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Brussels, Belgium, 135-150.

Badi, R., Bae, S., Moore, J. M., Meintanis, K., Zacchi, A., Hsieh, H., Shipman, F., & Marshall, C. C. (2006). Recognizing user interest and document value from reading and organizing activities in document triage. In *Proceedings of the 11th International Conference on Intelligent User Interfaces*, Sydney, Australia, 218-225.

Bae, S., Kim, D., Meintanis, K., Moore, J.M., Zacchi, A., Shipman, F.M., Hsieh, H., & Marshall, C.C. (2008). Supporting document triage via annotation-based visualizations. In *Proceedings of the American Society for Information Science and Technology, 45*(1), 1–16.

Bae, S., Kim, D., Meintanis, K., Moore, J.M., Zacchi, A., Shipman, F.M., Hsieh, H., & Marshall, C.C. (2010). Supporting document triage via annotation-based multi-application visualizations. In *Proceedings of the 10th Annual Joint Conference on Digital Libraries,* Gold Coast, QLD, Australia, 177-186

Baldonado, M.Q.W., & Winograd. T. (1997). SenseMaker: An information-exploration interface supporting the contextual evolution of a user's interest. In *Proceedings of the 15th SIGCHI Conference on Human Factors in Computing Systems*, Atlanta, GA, USA, 11–18.

Bolt, R. A. (1984). *The human interface*. Belmont, CA:Lifetime Learning Publications.

Card, S.K., Robertson, G.G., & York, W. (1996). The web-book and the web forager: An information workspace for the world-wide web. In *Proceedings of the 14th SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, BC, Canada, 112–117.

Chang, S.K. (1989). Visual reasoning for information retrieval from very large databases. In *the 2nd Annual IEEE Workshop on Visual Languages*, Rome, Italy, 1-6.

Delicious (2010). Social bookmarking. Retrieved 2010, from http://www.delicious.com/

diSessa, A.A.(1985).   A principled design for an integrated computational environment. *HCI, 1*(1), 1-47.

Donahue, J., & Widom, J. (1986). Whiteboards: A graphical database tool. *ACM Transactions on Information Systems*, *4*(1), 24-41.

Fowler R.H., Fowler W.A. L., & Wilson B.A. (1991). Integrating query thesaurus, and documents through a common visual representation. In *Proceedings of the 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Chicago, IL, USA, 142-151.

Furnas, G.W. (1986). Generalized fisheye views. In *Proceedings of the 4th SIGCHI Conference on Human Factors in Computing Systems*, Boston, MA, USA, 16–23.

Gheel. J, & Anderson. T, (2007). Activity metadata for enhancing web document retrieval. In *Proceedings of the 21st British HCI Group Annual Conference Workshop on Supporting Human Memory with Interactive Systems,* Lancaster, U. K.

Gheel. J. (1999) Data and metadata for finding and reminding. In *Proceedings of the 3rd International Conference on Information Visualisation,* London, U. K., p.446.

Girgensohn A., Shipman F., Chen F., & Wilcox L. (2010). DocuBrowse: Faceted searching, browsing, and recommendations in an enterprise context. In *Proceedings of the 15th International Conference on Intelligent User Interfaces*, Honk Kong, China,189-198.

Goldberg. A. (1984). *Smalltalk-80: The interactive programming environment.* Reading, MA:Addison-Wesley.

Hearst, M.A. (1995), Title bars: Visualization of term distribution information in full text information access. In *Proceedings of the 13th SIGCHI Conference on Human Factors in Computing Systems*, Denver, Colorado, USA, 59-66.

Hearst, M.A., & Chandu. K.(1997). Cat-a-Cone: An interactive interface for specifying searches and viewing retrieval results using a large category hierarchy. In *Proceedings of the 20th Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, Philadelphia, PA, USA, 246–255.

Henderson, D.A. Jr, & Card, S. (1986). Rooms: the use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface. *ACM Transactions, 5*(3), 211-243.

Kim D. & Shipman F. (2010). Interpretation and visualization of user history in a spatial hypertext system. In *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*, Toronto, Ontario, Canada, 255-264.

Lamping, J., Rao, R., & Pirolli, P. (1995). A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. Katz R. I., Mack R., Marks L., Rosson M. B., & Nielsen J.(Eds.). In *Proceedings of the 13th ACM SIGCHI Conference on Human Factors in Computing Systems*, Denver, Colorado, USA, 401–408.

Mackinlay, J. D., Card, S.K., & Robertson, G.G. (1991).   PerspectiveWall: Detail and context smoothly integrated. In *Proceedings of the 9th ACM SIGCHI Conference on Human Factors in Computing Systems*, New Orleans, Louisiana, USA, 173–176.

Mann, M.T. (1999). Visualization of WWW-search results. In *Proceedings of the 10th International Workshop on Database & Expert Systems Applications, IEEE Computer Society*, Florence, Italy, 264-268.

Marshall, C. C., & Shipman, F. (1997). Spatial hypertext and the practice of information triage. In *Proceedings of the 8th ACM Conference on Hypertext,* Southamptom, U.K., 124-133.

Marshall C. C., Shipman F., & Coombs J.H. (1994). VIKI: Spatial hypertext supporting emergent structure. *In Proceedings of the ACM European Conference on Hypermedia Technology*, Edinburgh, Scotland, U. K., 13-23.

Mukherjea, S., & Hara, Y. (1999). Visualizing world-wide web search engine results. In *Proceedings of the 3rd International Conference on Information Visualisation*, London, England, U. K., p.400.

PageRank (2010). Google toolbar API.   Retrieved 2010, from http://en.wikipedia.org/wiki/PageRank#Google_Toolbar

Pygmalion S. D. (1975). *A creative programming environment*. Doctoral Dissertation, Stanford University, Stanford, CA.

Rao, R., & Card, S.K. (1994). The table lens: Merging graphical and symbolic representations in an interactive focus+context visualization for tabular information. In *Proceedings of the 12th ACM SIGCHI Conference on Human Factors in Computing Systems*, Boston, MA, USA, 318–322.

Robertson , G.G. & Mackinlay , J.D. (1993).   The document lens. *In Proceedings of 6th Annual ACM Symposium on User Interface Software and Technology*, San Jose, CA, USA, 101–108.

Shipman, F., Hsieh, H., Airhart, R., Maloor, P., & Moore, J.M. (2001).   The visual knowledge builder: A second generation spatial hypertext. In *Proceedings of the 12th ACM Conference on Hypertext and Hypermedia*, Arhus, Denmark, 113-122.

Smith D. C., Irby, C., Kimball, R., Verplank, W. & Harslem, E. (1982).   Designing the star interface. *Byte 7*, 242-282.

SurfCanyon. (2010). Retrieved 2010, from http://www.surfcanyon.com

Sutherland, I. E. (1963). Sketchpad: A man-machine graphical communication system. In *Proceedings of the AFIPS Spring Joint Computer Conference*, Detroit, MI, USA, *23*, 329-346.

tf-idf (2010). Term frequency inverse document frequency.   Retrieved 2010, from http://en.wikipedia.org/wiki/TF_IDF

Veerasamy, A., & Navathe, S.B. (1995).   Querying, navigating and visualizing a digital library catalog. *In Proceedings of the 2nd International Conference on the Theory and Practice of Digital Libraries*, p.16.

Vicente, K. J., & Rasmussen, J. (1990). The ecology of human-machine systems II: Mediating 'direct perception' in complex work domains. *Ecological Psychology*, *2*(3), 207–249.

Web Archive (2010). Internet archive. Retrieved 2010, from http://www.archive.org/web/web.php

Wordle (2011). Beautiful word clouds. Retrieved 2010, from http://www.wordle.net

APPENDIX 1

HEURISTIC EVALUATION QUESTIONNAIRE

Questionnaire

Evaluation of a System Layer design for the visual knowledge builder

Subject ID: _____

Usefulness

1. It is useful:

         1   2   3   4   5   6   7
   Strongly disagree   Neutral   Strongly agree

2. It helps me be more effective:

         1   2   3   4   5   6   7
   Strongly disagree   Neutral   Strongly agree

3. It helps me be more productive:

         1   2   3   4   5   6   7
   Strongly disagree   Neutral   Strongly agree

4. It saves me time when I use it:

         1   2   3   4   5   6   7
   Strongly disagree   Neutral   Strongly agree

5. It meets my needs:

         1   2   3   4   5   6   7
   Strongly disagree   Neutral   Strongly agree

6. When will it be useful?

_____
_____
_____
_____
_____
_____
_____
_____

7. General comments about this part:

_____
_____
_____
_____
_____
_____
_____
_____

The above sets of questions were asked for each of the new features of the System Layer

to evaluate their usefulness.

Ease of use

1. It is easy to use:

        1   2   3   4   5   6   7
    Strongly disagree   Neutral   Strongly agree

2. It is simple to use:

        1   2   3   4   5   6   7
    Strongly disagree   Neutral   Strongly agree

3. It is flexible:

        1   2   3   4   5   6   7
    Strongly disagree   Neutral   Strongly agree

4. Using it is effortless:

        1   2   3   4   5   6   7
    Strongly disagree   Neutral   Strongly agree

5. I can use it without written instructions:

       1   2   3   4   5   6   7
Strongly disagree   Neutral   Strongly agree

6. It is user friendly:

       1   2   3   4   5   6   7
Strongly disagree   Neutral   Strongly agree

7. General comments about this part:

_____
_____
_____
_____
_____
_____
_____
_____

Ease of learning

1. I easily remember how to use it:

       1   2   3   4   5   6   7
Strongly disagree   Neutral   Strongly agree

2. It is easy to learn to use it:

       1   2   3   4   5   6   7
Strongly disagree   Neutral   Strongly agree

3. General comments about this part:

_____
_____
_____
_____
_____
_____
_____

Satisfaction

1. I am satisfied with it:

       1   2   3   4   5   6   7
Strongly disagree   Neutral   Strongly agree

2. I would recommend it:

       1   2   3   4   5   6   7
Strongly disagree   Neutral   Strongly agree

3. I feel I need to have it:

       1   2   3   4   5   6   7
Strongly disagree   Neutral   Strongly agree

4. General comments about this part:

_____
_____
_____
_____
_____
_____
_____
_____

List the most positive aspects of this design and features.

_____
_____
_____
_____
_____
_____
_____

List the most negative aspects of this design and features.

_____
_____
_____
_____
_____

APPENDIX 2

HEURISTIC EVALUATION COMMENTS AND SUGGESTIONS

Usefulness

Summary

When will it be useful?

Participant 1: finding related information rapidly

Participant 2: It will be useful if it were longer. I guess the clicking on the page and see the page itself requires less steps. Unless several documents will be open on the summary page at the same time, so it is possible to compare them.

Participant 3: There is a lot of data on the Internet, we need tools and algorithms to efficiently identify and categorize useful documents.

Participant 4: To get a quick look at the page.

Participant 5: Quickly find important information about my interesting.

Participant 6: It can be used for personalization, search, and important as crucial features.

Participant 7: I certainly think it is useful. When the page has loads of information it saves lot of time.

Participant 8: In a world of terrible search engines, the summary would be an incredible feature for me. The reason I say this is that, given the functionality of a search engine and the way it works, in an era of search engines, a user would be totally aware of the

input keywords he\she gives and the results obtained. So, summary wouldn't make a big difference.

Thus the summary is a great substitute for a search engine.

Participant 9: As Summary tab provided highlights of the main content being searched this would be necessary for user to get an idea to validate the result

Participant 10: It is always useful to check the summary before opening a document. The summary gives the user an idea about the possible content in the document and helps him decide to open or not the document. It improves the browse time and overall experience to the user.

General comments/suggestions about this part (Summary)

- Make is a little bit longer

- Apart from classifying documents, this can be extended to classifying emails in an inbox.

- It properly uses not only personal (implicit) preference, but also explicit features such as meta data.

- If summary is extended to touch on more parts of the Web Document, it could be used even better.

- May be the summary can be used in some computations (can be with other summaries) with a small factor, as it had a small percentage for a user to choose a result.

- This summary might also involve some frequently repeating words/phrases in the document apart from text. They can be calculated just based on term frequency in

the document after removing the stop words. (These will be different from IPM keywords and popular tags as they are just summary of the document).

IPM keywords

When will it be useful?

Participant 1: finding related documents understand what issues are there regarding the provided search terms.

Participant 2: It will be useful to know why the system is suggesting me some pages, to have an idea of the criteria.

Participant 3: To find out why a document has a specific color and the dominant keywords in it.

Participant 4: It gives you a pictorial assessment of the frequency of words in the content and helps in estimating relevance of the page to your search .

Participant 5: Visually, I could find easily that I want to find it. Awesome interface.

Participant 6: The more I use it, the better the system can satisfy me, I guess.

Participant 7: As this part picks the keywords user searches for frequently while browsing, this helps in classifying a page quickly into his interested or disinterested categories. This saves much time in picking the pages of our choice.

Participant 8: I find this incredibly useful in making me search engine friendly. If I didn't know how to use keywords in a search engine, by using this keywords feature, I can mentally map my interests against keywords. I can then better use the web to search for documents myself.

Participant 9: The words listed could give more importance to the idea being represented by the site. For example for a search item if for one result I see the words related to a place and another related to a person, it would help me to decide faster.

Participant 10:

- A frequent user of the VKB will find it more useful when the system suggestions the keywords/topics of discussion in the document.

- These personalized suggestions helps in helping the user in finding out the topics of discussions in the document looked from his/her perspective

- Helps the user in finding out new knowledge(keywords) that help in refining his search terms/criteria.

- It also helps the user in knowing what his interest trends have been in history looking at the interest terms and their magnitude.

General comments/suggestions about this part (IPM keywords)

- The better the algorithm to arrive at similarity, the more accurate the keywords will be.

- Some unnecessary words were shown.

- It needs better formatting.

- It would be nice if on click of the keyword, we are taken to the webpage with the words highlighted .This will help us glide through the page sooner for what we are looking for.

- How about animation effect or color?

- Some IPM keywords is not strongly relevant to the search results.

- Most keywords presented are very relevant and useful (except for some trivial keywords like http, www).

Popular tags

When will it be useful?

Participant 1: Understand other users' suggestion.

Participant 2: to have an idea of the content of the document by looking at the tags other people put on it

Participant 3: Useful to see the popular tags from a community based site like deli.cio.us

Participant 4: Helps to better categories the page based on views of a group of people.

Participant 5: This saved my time to search.

Participant 6: The tags are reliable key words in general.

Participant 7: useful when the page is tagged well(sensible as well as understandable to the user) and you know what tag you are searching for.

Participant 8: An excellent way of bookmarking, I can easily look for the correct content by eliminating documents on the tags. If these keywords can be integrated into the disk search feature of a PC's OS, the PC would be empowered with a mini search engine.

Participant 9: Info from entire sample space of users. Same as previous one. Can help to decide by associated tags to understand meaning behind the site

Participant 10: It helps the user in finding out what the rest of the world is thinking about the document and the favorite topics of interest related to the document. This also helps the user in getting more broader understanding of the document and generally helps in refining his search criteria resulting in more quickly finding the required information.

General comments/suggestions about this part (Popular tags)

- Showing how strong the tags would be good.

- Exploiting "delicious.com" is a good idea!

- The more the online resources are used to arrive at the tags, the better it is to make the tags effective and relevant.

- not sure, if this can be compared to the words listed by ipm..

Metadata

When will it be useful?

Participant 1: When i want to check the usefulness of the document.

Participant 2: to have an idea of the ranking and to help me evaluate the search results

Participant 3: To find Google's perception of the document reputation.

Participant 4: When viewing a higher ranked page it gives you a satisfaction of going through a more relevant page to your search.

Participant 5: It was useless to search on website. Maybe, this is for web developer or other industrial field's worker's needs. As a routine user, this menu is useless. I am not interesting about the specification about the size, number, and page rinks of the web site.

Participant 6: When I want to measure or estimate how the page is popular or well-known among people.

Participant 7: Page rank part of the Metadata is useful, especially because its definition in this context makes sense to a layman. Not sure how data size matters in most cases.

Participant 8: Information like PageRank may not make more sense than a keyword or tag. Only in case of exceptionally high page rank will it highlight the fact that this Web

Document is widely popular. Again the outgoing links will only give little information about the content or the relevance of the contents, in case of very high number of outgoing links.

Participant 9: Metadata like Creation date or modification date can give how old the data is. Page ranks can determine confidence level of the result

Participant 10: It might be useful to the user in finding out how important a document is (through page rank) before deciding to go ahead and read it as it saves time.

The size of the document information also helps the user in estimating the length of the document to read and if he is looking only for certain size(smaller/bigger) documents they can be filtered beforehand saving time.

General comments/suggestions about this part (Metadata)

- I didn't understand what PageRank value meant first.

- Doubtful, but will standardization of common meta data help..

- I think it might help to show if possible of the last updated time of the document. This can help the user in knowing if something has been updated in the document since he/she visited last time.

Contained links

When will it be useful?

Participant 1: Go to the related link directly without examining the document

Participant 2: It's nice to collect them and see how many link it has, or to see how many are out of the website

Participant 3: To find sites referred by the document and find their popularity.

Participant 4: Providing all links available in the page makes it easier to reference related links.

Participant 5: Broadly, this is useful to use to refer other links to find more information; however, it was hard to define what the best web site link is for me. Maybe, google searching would be better than this menu.

Participant 6: When I want to see what roles the search result plays as a hub, or what information it has, it will be useful.

Participant 7: I did not find this one to be critically useful. Especially because, links do not convey much meaning. If Title of the page was presented instead of the link that would have helped more I felt.

Participant 8: If I don't really find the contents very relevant for my interests, this feature opens up further ways of explicitly going through more Web Documents to build my interest profile

Participant 9: Faster navigation...

Participant 10: It might be useful many times in quickly finding the related links of a document. These links might contain something more valuable than the document itself so that user can quickly jump to more relevant documents saving time. (Though this requires some Meta information about the link)

General comments/suggestions about this part (Contained links)

- Showing the title for each web url would be good.

- The url could show strange documents so providing the usefulness based on the keywords would be helpful.

- Can include the PageRanks of the outlinks.

- It would be Ideal if the urls were actual hyperlinks to the page.

- In some contained links needs some interpretation or explanation or hyperlink to the corresponding webpage because I don't know what the pages the links indicates are.

- Instead of showing all links the user might be more interested if they could be filtered and ranked based on combination of some parameters like PageRank + frequency. Showing fewer might help in user finding more easily.

- Also allowing the user to quickly build a VKB collection out of these links might help in finding out more meta information about each of the links with a single click

- Showing more meta information with each link like heading and some summary terms etc. are most important I feel.

Web Archive

When will it be useful?

Participant 1: When I want to know if some part has been changed. Can check which parts have updated.

Participant 2: I think it can be useful when I'm looking for something that is not there anymore, or that I retrieved and now it is different from how I remembered.

Participant 3: To find the history of how the site has been changing.

Participant 4: By comparing the state of web page at two different points of time , we can ascertain if or not the page has evolved or been active . It is assuring to read or reference pages in a blog or forums which are active .

Participant 5: In terms of comparison of version history, this was very useful to find some changed information.

Participant 6: To trace the page history.

Participant 7: I believe, this is helpful when the content of the pages are changing frequently.

Participant 8: When watching how a socially collaborative webpage, say a wiki like wikipedia, evolves and how users have contributed, Web Archive will be very intuitive to understand how changes have happened. I consider this feature as the most visual and interesting of all.

Participant 9: Historical information can indicate activity and validity of information. Useful for news or modification of source code etc..

Participant 10: This sure helps in understanding how the current document or the content inside the document has evolved across time. This is more useful while researching on frequently changing topics and could help in easily pointing the changes and analyzing their causes quickly.

General comments/suggestions about this part (Web Archive)

- Showing brief summary of update through the thumbnail

- It will be nice to also archive snapshots of the users view of the pages.

- Good to have an option to view/compare archive from a particular day or period.

- Pages with dynamically changing content may not need historical picture of changes (news pages, scores) very often. However, for pages from sites like wikipedia, this feature can be very helpful!

General comments/suggestions about this part (Ease of use)

- Easy to use. Simple.

- It is hard to find the purpose of some of the buttons. It would be helpful if we had a status bar or some way of finding out the purpose of buttons.

- With text interface, it seems like that hard to learn at the first time. I needed more interactive and easier interface, GUI.

- Not complicated to use.

- Just clicking tabs or icons are all I had to do, so I could use that with just my intuition.

- It's quite easy to use and most of the information presented is easily understandable without needing documentation.

- And it's very easy to navigate across tabs and get the required information.

- More visual cues are required to guide user to know that such a layer exists apart from documenting that such a thing exists. For example, a simple side arrow bar on the User layer will hint the user that another layer exists behind the user layer.

- If this tool can be extended to work with say documents within a standalone PC's disk, it could be an awesome tool for effectively using the PC. I can easily find where my tax documents are, where my research papers are, etc

General comments/suggestions about this part (Ease of learning)

- It was simple.

  Several minutes explanations were enough.

- Components are simple, effective and distinct. So easy to learn overall.

- A small description, in few words, of each tab in a tooltip may help the user to

  better understand what each tab is for.

General comments/suggestions about this part (Satisfaction)

- If some other features are added, I expect the system to be improved more. :)

  Thanks.

- The color tagging combined with this extra data provided to user will make much

  difference in his productivity in searching through web.

- I would love it use it as a substitute for my delicious bookmarks. Also, I am

  thinking that I can get rid of my sticky notes if I start using this.

List the most positive aspects of this design and features

- I liked the Web Archive function and summary the possibility to see different

  features in different documents

- Enhance user experience while encountering new documents based on past

  browsing history. Can be extended to emails to automatically classify emails into

  different labels.

- Good program to collect large website and organize well to work my research or

  work. In terms of learning, it was simple and easier than other tools. Thank you

  for your work.

- Easy to navigate

- Good ideas using both explicit and implicit features.

- Very simple to use and easy to learn and has quite a few useful details.

- I find many more usefulness of entire tool and also the System Layer in specific

- Easy navigation between tabs

- Adds more value for my interests. Makes me understand my interests better

- Opens up ways of building my interests on unexplored related areas

- Faster search and personalized search. It can surely help to decide faster on the search results.

- I liked the one common advantage in all the features of how they help the user in finding a lot of information about the document even before opening it. It saves lot of time.

List the most negative aspects of this design and features

- Summary too short.

- UI can be more descriptive; can have a tooltip or a status bar.

- I needed GUI and more colorful background and texture to make me feel good and happy. Text was not enough to organize well. Do you have also searching function on your program. If not, how about making a searching function on your work. Thank you.

- Several IPM keywords not strongly relevant to search result.

- No explanation of contained links

- Some features may be more useful with minor modifications - Ex: Title of the page instead of links;

- Missing a few visual cues

- Metadata may not be very effective in most cases

- Concepts like online social bookmarking may overlap with these features, throwing the usefulness of this tool into a risk

- Instead of showing different types of information in different tabs, it would help to consolidate some of them as a single category, like for example summary can also have IPM tags and popular tags ant one place.

- Also more customization could be added in latter stages.

- A suggestion (not negative): Also could be research whether while providing this different information if spatial concepts can also be used within a document too?

VITA

Name: Arun Bharath Gomathinayagam

Address: Department of Computer Science and Engineering
                Texas A&M University
                301 Harvey R. Bright Building
                College Station, TX 77843-3112

Email Address: arunbharath@gmail.com

Education: B. Tech, Information Technology, PSG College of Technology, 2005
                M.S., Computer Science, Texas A&M University, 2011

Research: Spatial Hypertext, Computer Human Interaction and Computer
                Supported Collaborative Work.