Investigating Aspects of Visual Clustering in the Organization of Personal Document Collections

Hoda Badesh
Faculty of Information Technology
Misurata University
Misurata, Libya
hoda.badesh@gmail.com

Anwar Alhenshiri
Faculty of Information Technology
Misurata University
Misurata, Libya
alhenshiri@gmail.com

Jamie Blustein
Faculty of Computer Science
Dalhousie University
Halifax, NS, Canada
jamietab@cs.dal.ca

Evangelos Milios
Faculty of Computer Science
Dalhousie University
Halifax, NS, Canada
eem@cs.dal.ca

***Abstract*—Organizing personal collections of digital documents can be frustrating for two main reasons. First, the effort required to work with the folder system on personal computers and the possible misplacement and loss of documents. Second, the lack of effective organization and management tools for personal collections of digital documents. The research in this paper investigated specific visualization and clustering features intended for organizing collections of documents and built in a prototype interface that was compared to a baseline interface from previous research. The results showed that those features helped users with: 1) the initial classification of documents into clusters during the supervised stage; 2) the modification of clusters; 3) the cluster labelling process; 4) the presentation of the final set of organized documents; 5) the efficiency of the organization process, and 6) achieving better accuracy in the clusters created for organizing the documents.**

Keywords- information organization, management, retrieval, clustering, visualization, human factors.

I. INTRODUCTION

Personal documents grow in size and number rapidly. In the current state, desktop documents can be organized either manually in folder hierarchies or using special software such as: OpenText[15], IBM's Document Manager, and Google Desktop[16], which has been discontinued). Manual organization can be very demanding since desktop computers may involve large collections of documents. Every type of software has its advantages and disadvantages. For instance, Google desktop presented its search results in a list provided from searching the index of keywords Google built from the user documents. This type of presentation may require the user to go through large lists of result hits, formulate several queries, and eventually may or may not find the intended document.

When the pile of documents on a user's desktop grows extensively, organizing those documents into folders may become very time consuming. The use of software that presents lists of results may also be very ineffective. The use of clustering for organizing user desktop documents has had little consideration. User interfaces for assisting users with organizing their documents using aspects of clustering and visualization have not been thoroughly investigated.

Clustering is grouping together documents of the same type, genre, topic, etc. A categorization scheme has to be defined prior to applying clustering. Topical clustering and genre clustering have been investigated [5][15]. The use of clustering in document presentation has been investigated for desktop retrieval as well as web retrieval [1][14]. Clustering makes use of overviews of documents for conveying the different topics or genres covered in the document collection.

Visualization can help the presentation of multiple features of search results [1][2]. Document features such as its size, last update, and type can be visualized. Features of the collection as a whole can also be visualized by showing documents of the same type connected or by showing documents with similar content under one category. Such visual clustering combines the benefits of visualization and clustering. Adding clustering and visualization to the presentation of search results can help users organize large collections of documents and find results more effectively and efficiently.

There are several problems associated with managing and organizing personal documents on desktop computers. The following summarizes those problems:

1. The size of the collection of documents on computers of personal nature grows very rapidly as users keep using their machines.
2. Manual organization of documents on desktop computers necessitates the use of folder structure which may result in:
3. Excessive time consumption in the case of large collections.
4. Losing documents due to the complex structures and the difficulties associated with manually searching those structures.
5. Organization tools may drive the user away for one or more of five reasons: visibility, integration, co-adoption, scalability, and return to investment [13].
6. Search using desktop tools has problems associated with the presentation of the search results and the interaction with the user.

The research discussed in this paper attempts to answer the following questions:

1- What is the effectiveness of using three options of document views (abstract, text cloud, full content) on how users classify their documents for organization?
2- What is the effectiveness of presenting the initial clusters during the classification process as bubbles containing glyphs of documents inside each

 ISBN: 978-1-61208-361-2

corresponding cluster with different modification capabilities?

3- What is the effectiveness of having different views of clusters, as a list of cluster labels and as labelled bubbles?
4- What is the effectiveness of presenting the final set of documents clustered and organized in bubbles representing topics with their documents represented as glyphs?

The features indicated in the questions above were investigated in a prototype interface called the Bubbles Interface. The prototype was essentially intended to investigate these particular visualization features (also shown in Table 1) in improving the organization of collections of personal documents using clustering. To investigate the usefulness of those features, the interface was compared to the Pie Interface, another project developed in [9] for the same purpose. The results of the research showed that the new interface had a better layout and assisted its users with the initial classification of documents. Modifying and labelling clusters was also enhanced using the new interface. The interface improved the final organization of the documents by improving the accuracy of the clusters created.

The remainder of this paper is organized as follows. Section 2 discusses related work. Section 3 illustrates the study conducted in this research. Section 4 provides a detailed discussion of the results. The paper is concluded in Section 5.

II. RELATED WORK

Managing and organizing information has been explored in different directions. Knoll et al. [14] investigated how users view and manage desktop information in general. Jones et al. [13] investigated important reasons behind giving up on certain personal information management tools. The strategies users follow to manage web information in order to be able to relocate and reuse information previously found are discussed in [12]. Their work showed that users follow different keeping strategies to re-find and compare information later. The variety of managing and organizing strategies for personal information can be attributed to the fact that current tools lack important reminding, integration, and organization schemes [7].

Jones et al. [14] found that users abandon the use of an information management tool for one or more of five closely related reasons: visibility, integration, co-adoption, scalability, and return on investment. Jones [11] reviewed research in support of a more general preference for *way finding* methods that depend on a sense of digital location vs. direct search as the primary means for access to personal information [6]. Bergman and Nachmias [4] indicated that direct search becomes the user choice for retrieving personal information after attempts for search by navigation fail.

Jaballah [10] designed a desktop personal library manager to overcome the problems associated with the use of folder-based organization schemes. Users could browse and search their personal collections of documents by the document type, title, filename, date of modification, and so on. The interface was evaluated using a pilot study (two experts) followed by a learnability study and final diary study (6 participants). The results showed that even with the prototype's ability to harvest metadata about the files in the collection, the users preferred the standard folder system. They reported that some actions on the prototype were difficult and that users spent most of the time trying to familiarize themselves with the interface.

To further emphasize the value of visual access to information for managing and organizing personal collections, Bauer [3] built an interface intended to arrange piles of images or PDF documents in portraits. Each PDF file in the portrait is shown as one page containing images and parts of the text in the documents. Images are shown in their own piles. The closer the image to the user, the larger the size of the document is. The prototype allowed interactions with collections of documents to be logged over long periods. The prototype was not evaluated and it was expected to improve the user's experience with managing piles of personal documents and images.

Civan et al. [6] compared the user behavior for organizing information using folders and using labels (tags). For the purpose of the comparison, Gmail, which is Google's email service, and Hotmail, which is Microsoft's email service, were selected. Users organized their e-mail messages using different methods in the two systems. Gmail's users labeled or tagged their messages; Hotmail users put messages into folders. The two approaches were compared with respect to: "retrieval performance, evolution in mappings between articles and folders/labels over time, and limitations to fully express one's internal conceptualization" [6]. No clear winner was identified between tagging and placing. The study concluded that "better support for information organization may need to go well beyond folders and tags or their artful combination" [6].

Managing information is concerned with how people store, organize, and re-find information [8]. Information management systems are methods by which users find, categorize, and re-find information on daily basis. Research has considered personal information management. However, there is further need for investigating organizing and finding information in cases where the personal collection of documents grows extensively and when standard folder-based organization becomes overwhelmingly demanding.

III. RESEARCH STUDY

The study discussed in this paper compared two interfaces, the Pie Interface from the work in [9] and the Bubbles Interface designed for the purpose of this study. The Pie Interface was selected based on the results of a previous study that showed some drawbacks in the prototype during the evaluation. The Bubble Interface was designed and compared to the Pie Interface for evaluating the features embedded in the Bubble Interface to overcome difficulties users encountered with the Pie Interface in the previous study as discussed in [9]. The interfaces are briefly described as follows:

 ISBN: 978-1-61208-361-2

Figure 1. The Pie Interface.

A. *The Pie Interface*

The Pie Interface is divided into four sections: 1) the supervision panel, 2) the un-assigned document view, 3) the cluster view, and 4) the labeled document view. They are shown in Figure 1.

B. *The Bubble Interface*

This interface was designed to allow users to organize their personal collections of documents based on clustering and using aspects of visualization in both the classification stage, which is the supervised portion of the process, and the final presentation stage of the organization process. The Bubbles Interface (shown in Figure 2) was designed to overcome several disadvantages in the Pie Interface.

C. *Study Design and Population*

The study design was complete factorial and counterbalanced. It accounted for the possible effects of *order* using two conditions in a *within-subject* design. The possible main effect of the independent variable (the interfaces) was controlled by randomly selecting with which interface the participant started.

 ISBN: 978-1-61208-361-2

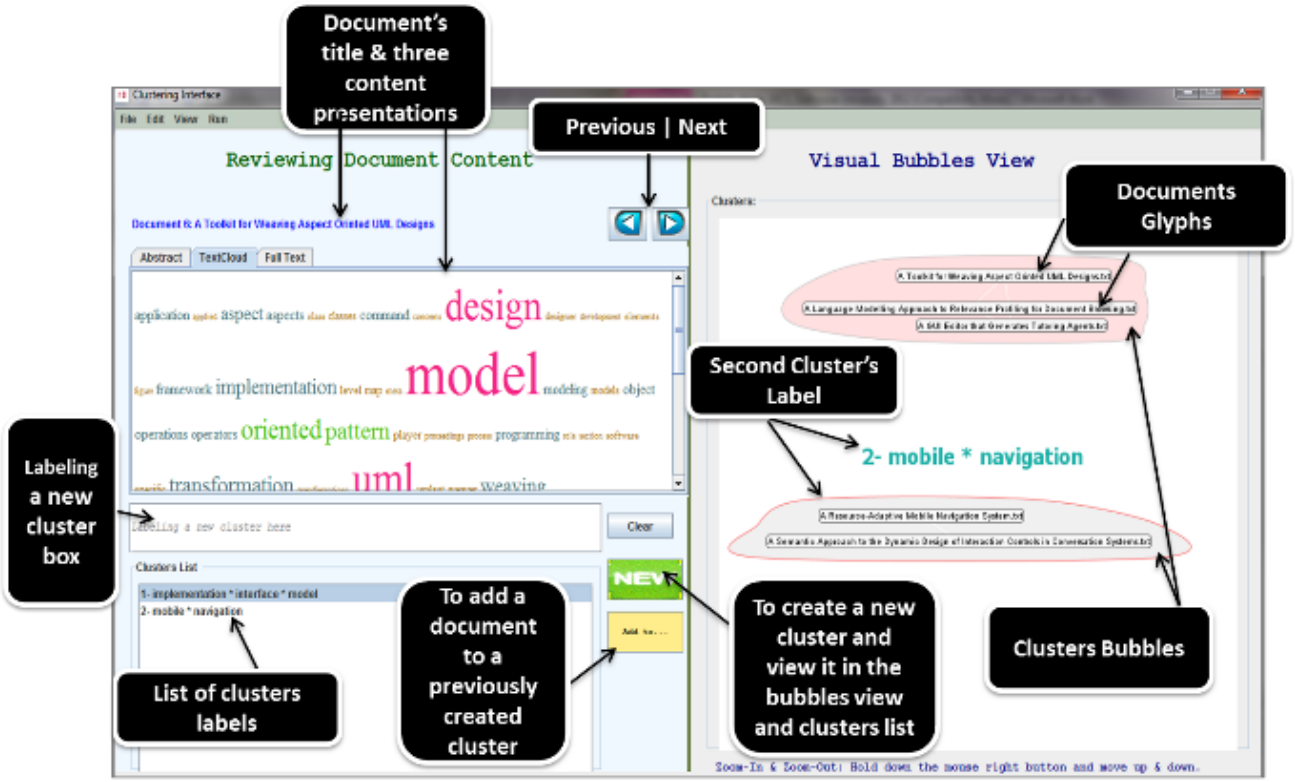

Figure 2. The Bubble Interface.

Ten participants—all computer science students—took part in this study. The small sample was meant to provide evidence of effectiveness of the prototype for further studies. The participants, even though are few, they represent early adopters. Of the participants, eight were males and only two were females. The ages of two participants were between 18 and 22. The ages of the remaining eight participants were between 23 and 30. All participants were graduate students.

D. Study Methodology

Every participant was given 30 documents (randomly selected from the collection used in the previous experiment) from which they could select 12 documents as seeds to clusters (1–12 clusters). They were given 15 minutes to classify the 12 documents into initial clusters. This was the supervision stage. The ten participants were split into two teams (Team 1 consisted of four participants while Team 2 consisted of six participants). The two teams met on two different days. On the first day of the study, each team had a meeting and the evaluation was completed as follows:

1. The team was divided into two groups (Group 1 and Group 2).
2. Each group was given a training session (approximately 5 minutes) on how to work on each interface.
3. Group 1 started working on the Bubbles Interface while Group 2 started working on the Pie Interface.
4. The participants were given the 30 documents used in the study two days ahead to familiarize themselves with the collection.
5. Every participant was asked to classify 12 documents from the collection of 30 documents into any number of clusters (1-12 clusters). After completing the classification process, the interface called the underlying clustering algorithm used in the work discussed in [9] and the remaining 18 documents were assigned by the algorithm to complete the clustering stage.
6. Every participant was asked to evaluate the clustering process by deciding whether or not each of the documents was assigned to the correct cluster from the participant's point of view.
7. Every participant was asked to complete a post-testing questionnaire about the interface they used.
8. The groups were then switched to follow the same steps 5 through 8 as described above.
9. A focus group discussion took place after completing the task on both interfaces.

 ISBN: 978-1-61208-361-2

TABLE I. A COMPARISON OF THE FEATURES ON THE PIE AND BUBBLE INTERFACES

No.	Features	Pie Interface	Bubble Interface
1.	Document representation	In a circle with a document ID	As a document title with a document index
2.	Mechanism of showing documents	Automatically after classifying the previous document	Using "*Previous*	*Next"* buttons
3.	Permanent document content view(s)	Plain text cloud + whole content	Abstract only + colorful text cloud + full text
4.	Other document content view	None	PDF format in a new window
5.	Creating clusters mechanism	Drag and drop a document into the "*New Cluster*" sector to create a new sector (cluster) containing a yellow stripe (document)	Click the "*New"* button. The new cluster label will be added to the Clusters List. A new bubble (cluster) with a glyph (document) will appear in the Visual Bubbles View
6.	The case of creating a cluster without a label	Although it is incorrect, the interface allows users to do so.	An error massage will pop up asking the user to create a label first.
7.	Visual view of clusters	a) Pie chart presentation b) Stripes (documents) within a sector (cluster) c) Not zoom-able	d) Visual bubbles presentation e) Glyphs (documents) within a bubble (cluster) f) Zooming in and out, and moving the bubbles around
8.	Viewing one cluster at a time	Not applicable	Allowed
9.	Skipping document(s)	Users are allowed to do one of two things to a document. 1) Assign it to a cluster. 2) Send it to the "*Trash"* sector (will not be considered in the clustering phase).	Allowed by hitting the "*Next"* button

A background questionnaire was used to gather demographic data about the participants. It was also used to collect information about the size of the participants' personal collections of documents and any tools they use to organize their documents.

The study was meant to evaluate the effectiveness, efficiency, and enjoyment of each interface and compare the two interfaces. The efficiency was measured using the time and the number of mouse clicks needed to complete the study. The perceived effectiveness and the engagement of the interfaces were measured through the data accumulated in the questionnaires and the accuracy of the clustering process.

The design of the experiment in [9] influenced the design of the current experiment in many ways. First, both studies used the same collection of documents. Second, the current study gave users the documents in advance since they were unhappy about the time they took to familiarize themselves with the documents in the study illustrated in [9]. The design of the Bubbles Interface attempted to change the visualization used in the Pie Interface and provide more interaction and display of content features, as seen in Table 1.

E. Study Results

1. Efficiency Result

The number of mouse clicks (left, right, and middle) during the study was logged. The number of mouse clicks on the Pie Interface was 301.3 on average (SD=66.6). In the case of using the Bubbles Interface, the average number of mouse clicks was 208.5 (SD=142.41). A two-sample-for-means z-test showed that no significant difference existed between the number of mouse clicks on the Bubbles Interface and the number of mouse clicks on the Pie Interface ($\boldsymbol{z}$ = -1.86, p = 0.06). However, by removing the outliers in the case of the Bubbles Interface, the difference became significant ($\boldsymbol{z}$ = -6.22, p<0.0001).

2. Effectiveness Result

To measure the effectiveness of the interfaces and compare the Bubbles Interface to the Pie Interface, every participant was asked to evaluate the accuracy of the final clustering of the 30 documents used. Every participant was asked to determine which documents were assigned to the correct clusters and which documents were assigned to the incorrect cluster based on the cluster topic built by the participant. The two-samples-for-means z-test results ($\boldsymbol{z}$ = -

 ISBN: 978-1-61208-361-2

2.93, p <0.003) indicated that there was a significant difference between the two interfaces with respect to the number of documents accurately clustered as perceived by the participants.

3. Enjoyment Result

The study used a post-task questionnaire for each interface after the user completed the task. Each questionnaire had 16 five-point Likert-scale questions that measured engagement factors considered in the study. The questions used involved the option of 'other' in most cases so that the user could provide different answers from the choices given. In all of the questions that used Likert-scales, the neutral case (i.e. the answer of 'not sure') was ignored from the analysis.

The first and second choices of the 5-point Likert-scale were merged and considered as one choice. The same procedure was followed with the fourth and fifth choices. The data was evaluated using the *z-test* (Downy et al., 2004) for comparing two proportions (equivalent to *Chi Square*). The following discussion goes through the results in each individual case measuring the engagement of the interfaces.

a. **How easy was the selection of documents for each cluster?** Nine participants chose 'easy' and 'very easy' for the Bubbles Interface, while only three participants found the Pie Interface to be 'easy' with regard to selecting documents for each cluster. The difference between the two proportions of participants (9/10 and 3/10) was significant ($\mathbf{z} = 2.739$, $p < 0.007$).

b. **How effective (helpful and useful) did you find creating labels for a new cluster?** On the Bubbles Interface, eight participants (8/10) indicated that creating cluster labels was 'effective'. The remaining two participants selected the neutral choice 'not sure' on the Likert-scale. On the Pie Interface, five participants chose 'effective' while three participants selected the 'not effective' choice. The difference between the two proportions of participants who considered the labelling feature on either interface as effective (8/10 and 5/10) was not significant ($\mathbf{z} = 1.41$, $p = 0.16$).

c. **How easy was modifying a cluster to add or remove documents?** On the Bubbles Interface, 70% of the participants (7/10) found it easy to modify clusters created during the supervision stage. Two participants indicated that it was difficult while the remaining one selected the neutral choice 'not sure'. On the Pie Interface, eight participants (8/10) found modifying clusters to be easy. One participant found it to be difficult while the remaining one was 'not sure'. The difference between the proportions of participants was not significant ($\mathbf{z} = -0.52$, $p = 0.60$).

d. **How clear did you find the view of your selected documents in the initial clusters?** On the Bubbles Interface and during the supervision stage, six participants (6/10) liked the clear presentation of their initial clusters. Two participants indicated that it was not clear while the rest selected the neutral choice 'not sure'. During the supervision stage on the Pie Interface, five participants liked the clear presentation of their initial clusters. Three participants found it unclear while two participants selected the 'not sure' choice. The difference between the proportion of participants who found the presentation of the initial clusters clear on either interface was not significant ($\mathbf{z} = 0.45$, $p = 0.56$).

e. **How helpful and effective did you find the final view of the clusters created by the system?** On the Bubbles Interface, four participants (4/10) found the final presentation of clusters to be helpful and effective. Three participants (3/10) indicated that it was neither helpful nor effective because of the overlapping of the documents' names while the three remaining participants (3/10) selected the neutral choice 'not sure'. On the Pie Interface, four participants (4/10) found the final presentation of the clusters to be helpful and effective. Four participants (4/10) considered it neither helpful nor effective while two participants (2/10) were 'not sure'. The difference between the proportions of participants who found the final presentation of the clusters helpful and effective on the Bubbles Interface and those who found it helpful and effective on the Pie Interface was not significant ($\mathbf{z} = 0$, $p = 0.99$).

f. **How do you rate the presentation of elements on the interface?** All participants (10/10) rated the presentation of elements on the Bubbles Interface as effective. Four participants (4/10) rated the presentation on the Pie Interface as effective while four participants rated it as not effective. There was a significant difference between the proportions of participants who found the presentation of elements on the Bubbles Interface to be effective and those who found the presentation of the elements on the Pie Interface to be effective ($\mathbf{z} = 2.93$, $p < 0.003$).

g. **How do you rate the positioning of the document view and cluster view on the screen?** The positioning of the document view and cluster view on the Bubbles Interface were considered effective by 70% of the participants (7/10). Two participants rated the views as not effective while only one participant selected the 'not sure' choice. On the Pie Interface, the positioning of the document view and cluster view were considered as effective by three participants (3/10). Four participants (4/10) rated the view as not effective and the remaining three participants (3/10) selected the 'not sure' choice. There was a significant difference between the proportions of participants who rated the positioning of the document view and cluster view on the Bubbles Interface as effective and those who rated the positioning of the document view and cluster view on the Pie Interface as effective ($\mathbf{z} = 2.25$, $p < 0.02$).

h. **How easy was it to undo actions on the interface?** On the Bubbles Interface, eight participants (8/10) rated the

ability to reverse actions as easy. One of the remaining two participants rated it as difficult and the other one selected the 'not sure' choice. On the Pie Interface, three participants (3/10) rated the ability to reverse actions as easy while three other participants (3/10) rated it as difficult. The remaining four selected the neutral choice of 'not sure'. The difference between the two proportions of participants who found reversing actions to be easy on either interface was significant ($z = 2.25$, $p < 0.02$).

i. **Was the feedback from the interface helpful to you?** The feedback from the Bubbles Interface was considered as clear and helpful by eight participants (8/10), not clear or helpful by one participant (1/10), and not applicable by one participant (1/10). The feedback from the Pie Interface was considered as clear and helpful by only three participants (3/10), not clear or helpful by two participants (2/10), and not applicable by one participant (5/10).There was a significant difference between the proportions of participants who found the feedback from the Bubbles Interface as clear and helpful and those who found the feedback from the Pie Interface as clear and helpful ($z = 2.25$, $p < 0.02$).

j. **How helpful and effective do you think the interface will be with organizing your collection of documents?** Seven users (7/10) predicted that the Bubbles Interface will be helpful and effective with organizing their own collections of documents. Two participants (2/10) anticipated that it will neither be helpful nor effective. Two participants predicted that the Pie Interface will be helpful and effective with organizing their own collections of documents. Four participants (4/10) anticipated that it will neither be helpful nor effective. There was a significant difference between the proportions of participants who expected the Bubbles Interface to be helpful and effective and those who expected the Pie Interface to be helpful and effective ($z = 2.24$, $p < 0.02$).

F. Study Limitations

The study had volunteers who were computer science students. The population of the study was very limited with regard to the number of participants involved due to limited resources. The number of documents used in the experiment was also limited because of the time required to manage more documents and investigate the effectiveness of the final clustering. The accuracy of clustering was manually examined which would have required more time and funding if more documents had been used.

IV. DISCUSSION

The study showed that users worked more efficiently on the Bubbles Interface than they did on the Pie Interface. The Bubbles Interface required significantly fewer mouse clicks by the user than the Pie Interface to complete the same task. However, there was no significant difference between the times needed to complete the task on the Bubbles Interface and the times needed by users to complete the same task on the Pie Interface.

Performing more clicks on the Pie Interface can be attributed to the user's need for very frequent scrolling in order to see the document content. This kind of scrolling was not needed as frequently on the Bubbles Interface. The reason for completing the tasks on both interfaces with no significant difference in the time needed can be attributed either to the nature of the task itself or to other factors that were not measured in the study.

Users achieved higher clustering accuracy with the Bubbles Interface than they did with the Pie Interface. One participant indicated that "*navigation among the document content views was much easier with the Bubbles Interface*". The Bubbles Interface may have helped users with assigning the appropriate documents together to represent a topic (cluster). It may have also helped users with identifying the documents in each cluster in the final results. The labelling process on the Bubbles Interface may have also helped with identifying the accurate topic of both the documents during the supervised classification stage and the clusters during the final presentation stage. One participant mentioned that "*I did very well in assigning documents into correct clusters."*

Several engagement factors have been addressed in the study. For example, the difference between the number of participants who found the process of selecting documents for clusters to be easy on the Bubbles Interface and those who found it easy on the Pie Interface was significant. This may indicate that the approach that was used to show the document content to the user was more effective on the Bubbles Interface. It may also indicate that users found it easier to perceive the cluster content and see where the new document belonged during the supervised initial classification.

Users also found the presentation of elements on the Bubbles Interface to be more effective than the presentation of elements on the Pie Interface. Users commented that the layout was intuitive and easy to understand and that no confusion or frustration was caused with the organization of the Bubbles Interface elements. During the group discussion, one participant stated "*I had some difficulties viewing the document content both with the text cloud and the whole content view. The area customized for displaying the content was not sufficient. It should be larger on the Pie Interface.*" Another participant indicated "*I got really lost with the Pie Interface because I always forget how to review the document and cluster content.*"

The positioning of the document view and cluster view on the display was considered effective and helpful by significantly more users of the Bubbles Interface than users of the Pie Interface. The participants reported that they "*found the interaction with the Bubbles Interface easier because of the nice layout that was easy to understand.*"

The feedback given by the interfaces was different. Significantly more users favoured the feedback given by the Bubbles Interface. For example, all the messages given by the Bubbles Interface were clear and were given to serve many purposes. However, the only feedback message given to the user while using the Pie Interface was the delete

 ISBN: 978-1-61208-361-2

conformation message when the user attempted to delete a cluster or a document. The users stated that the feedback of the Bubbles Interface was more helpful and reduced the need for asking the researcher questions to clarify the reactions of the interface. Two participants stated that they liked "*to have something on the interface indicating how many documents have already been classified and how many remain.*"

Users favoured the Bubbles Interface for future use with organizing their collections. Moreover, participants preferred the categorization of documents in the final results provided by the Bubbles Interface over the categorization provided by the Pie Interface. They indicated that in the case of the Pie Interface, there was little information about the documents in each cluster. The interaction with the interface to obtain more information about the clusters and the documents was hard.

Even though the Bubbles Interface has promising features in the organization of documents, it may have issues of clutter with very large collections. Different design parameters may need to be adjusted such as the glyph size for the document and the size of the bubble representing the cluster. The quality of clustering of a large collection of documents can be evaluated in the case of using the Bubbles Interface by evaluating the seeds selected for the clusters. It will be almost impossible (very time consuming) to ask users in a laboratory experiment to measure the accuracy of the final results in the case of very large collections of documents. However, the seeds chosen by the users to be given to the clustering algorithms can be evaluated by comparison to a ground truth.

There are several guidelines that can be drawn from the findings of the study. First, visualization through the use of intuitive bubble clusters would assist users in isolating clusters to locate documents in retrieval-based interfaces. Second, the use of labeled bubbles representing documents within clusters eases the process of identifying documents within clusters. Third, the interactive clustering in which changes are applied immediately to the visual view of clusters during the classification stage makes organization more effective. Fourth, providing different views of clusters may help in allowing users to continuously modify the groups of documents assigned to clusters according to changes in the observed topics. Finally, the use of intuitive clustering such that of the bubble interface would improve the user judgment of the documents assigned to each cluster.

V. CONCLUSION AND FUTURE WORK

The investigation compared specific features in a prototype interface against a baseline interface from previous research [9]. The results of the investigation showed that the new prototype interface had a better layout and helped users with: 1) the initial classification of documents into clusters during the supervised stage; 2) the modification of clusters; 3) the cluster labelling process; 4) the presentation of the final set of organized documents; 5) the efficiency of the organization process, and 6) the actual accuracy of the cluster for the organization process.

Further work will focus on the use of visualization and clustering parameters for more effective retrieval of documents in personal and even larger collections of documents. Studies will focus on improving the current prototype to provide efficient environment for organizing and managing in addition to retrieving documents.

REFERENCES

[1] A. Alhenshiri, and J. Blustein,"Exploring visualization in web information retrieval," International Journal for Internet Technology and Secured Transactions, vol. 3, issue 3, July 2011, pp. 320-330.

[2] H Badesh, and J. Blustein, "VDMs for Finding and Re-finding Web Search Results," iConference, Feb. 2012, pp. 419-420, Toronto, ON, Canada: ACM.

[3] D. F. Bauer, "Spatial Tools for Managing Personal Information Collections," HICSS2005, 2005, pp. 104-106, Hawaii, USA: IEEE Computer Society.

[4] O. R. Bergman, and R. Nachmias, "The project fragmentation problem in personal information," SIGCHI Conference on Human Factors in Computing Systems, April 2006, pp. 271-274, Montréal, QC, Canada.

[5] C. Carpineto, S. Osiński, G. Romano, and D. Weiss, "A Survey of Web Clustering Engines," vol. 41, issue 3, ACM Computing Surveys, July 2009.

[6] A. Civan, W. Jones, P. Klasnja, and H. Bruce, "Better to organize personal information by folders or by tags?: The devil is in the details," Journal of the American Society for Information Science and Technology, vol 45, issue 1, 2008, pp. 1-13.

[7] E. Cutrell, S. Dumais, and J. Teevan, "Searching to Eliminate Personal Information Management," In Communications of the ACM (Special Issue: Personal Information Management), Jan. 2006, pp. 58-64.

[8] D. Elsweiler, and I. Ruthavan, "Towards Task-based Personal Information Management Evaluations," the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2007, pp. 23-30, Amsterdam, The Netherlands: ACM.

[9] Y. Hu, E. Milios, J. Blustein, and S. Liu, "Personalized Document Clustering with Dual Supervision," the 12th ACM Symposium on Document Engineering, 2012, pp. 161-170, Paris, France: ACM.

[10] I. C. Jaballah, "Managing Personal Documents with a Digital Library," 9th European Conference, Research and Advanced Technology for Digital Libraries, 2005, pp. 18-23, Vienna, Austria.

[11] W. Jones, "Keeping Found Things Found: The Study and Practice of Personal Information Management," 2007, San Francisco, CA, USA: Morgan Kaufmann Publishers.

[12] W. Jones, H. Bruce, and S. Dumais, "How do People Get Back to Information on the Web? How Can They Do It Better? 9th IFIP TC13 International Conference on Human-Computer Interaction, 2003, Zurich, Switzerland.

[13] E. Jones, H. Bruce, P. Klasnja, and W. Jones, "I Give Up! Five Factors that Contribute to the Abandonment of Information Management Strategies. 68th Annual Meeting of the American Society for Information Science and Technology (ASIST 2008). 2008, Columbus, OH.

[14] S. H. Knoll, A. Hoff, D. Fisher, S. Dumais, and E. Cutrel, "Viewing Personal Data Over Time," CHI 2009 Workshop on Interacting with Temporal Data, 2009, pp. 1-4, Boston, MA, USA.

[15] M. Santini, and S. Sharoff, "Web Genre Benchmark under Construction," Language Technology and Computational Linguistics (JLCL), vol. 25, issue 1, 2009.

[16] OpenText, http://www.opentext.com

[17] Google Desktop, http://desktop.google.com

 ISBN: 978-1-61208-361-2