A network-based visualization tool for navigating news readings

Hsieh Ji-Lung [1 +], Shen Pai-Chen [1] and Yang Li-Wei [2]

[1] Graduate Institute of Library and Information Studies, National Taiwan Normal University, Taiwan, R.O.C

[2] Department of Information Management, National Taiwan Normal University, Taiwan, R.O.C.

Abstract. To assist users in news browsing, we built up a dynamic network visualization tool composed of visualized concept maps, trends of keywords, and interactive browsing paths, and related news contents. Users can drag and click on terms to expand related ones, traverse or alter the browsing path, and explore the concept network from different points of view. Three layout methods of visualizing concept maps were implemented to list and cluster concepts for different purposes. A simple user study was conducted to evaluate the usefulness and usability of our tool. The results of the user study showed both positive and negative opinions from our participants. They agreed that network-based visualizer provides related terms for the next browsing, and the interactive browsing path can help them control the set of related news. However, some of the participants indicated that the visualized features were too complex and even confusing for use.

Keywords: Information Visualization, News Visualization, Graph Drawing

1. Introduction

The amount of information grows rapidly with the development of Internet, and online news is no exception. Daily news readers are now suffering the "data smog," as described by David Shenk in 1997 [1]. Internet provides News agencies a publishing environment that is rapid, low-cost, 24hour, peer-produced, and easy-to-exchange. However, Internet also causes some information redundancy and reliability problems to daily readers. Readers can access the online news sites quickly, but they now spend more time reading what they want or all the news.

News reading behaviors can be casual, which is similar to the berry-picking process according to Bates [2]. Users often read news for casual "encountering" of interested events, not for "finding" or "seeking" something purposeful. However, current online news is presented in the form of separate pages containing several long lists of news titles and links, most of which are not centralized and well-organized as in the paper-based newspapers for an overview. The fact that readers must search and move in/out the hierarchy of links made it more difficult for them to encounter interested news. Besides, there are thousands of online news from either news agencies who provide both paper-based and online versions or news portals that aim to gather news together. In this case, the problem of data redundancy is not only existent but also worse due to the relative easiness to reproduce online news by "forwarding" and "copying."

The previous studies have developed auto-summarization and auto-clustering tools for users to retrieve or reorganize news efficiently. For example, Chen, et al. [3] proposed an ontology model to solve the problem of news redundancy. On the other hand, due to the increasing computing power of current personal computers (PCs), more and more services begin to provide information visualization interface to navigate users' browsing process. There have been several visualization tools for news browsing or monitoring [4]. For instance, *Google Wonder Wheel*, a navigating tool with a star layout, provides 5 or 6 related terms for users' browsing of the next retrieval. In terms of monitoring, *Blog Keyword Visualizer* and *Google Insight Search* are examples of dynamic visualization tools which focus on displaying associations among blog keywords and trends of related news respectively. Furthermore, the function of [5]'s dynamic social network

---

[+] Corresponding author. Tel.: +886277345436; fax: +886223620951.
  E-mail address: jirlong@gmail.com.

was not centered on the news content, reading, and news trend, but designed for discovering humans' relations when their names appear in the same news articles.

In recent years, network-based visualization which represents information by network graphs has been applied to users' daily internet surfing for searching, browsing, and organizing information [6]. In the network graphs, nodes are used to represent concepts or information, while links stand for associations (i.e., co-occurrence) between each pair of concepts. Most of the network visualization tools put their emphasis on how to better externalize the structure of information organization. The methods of these tools include clustering related news or listing related keywords such as a tag list or tag cloud in most of the social bookmark services. Our goal of the study, on the contrary, is to provide an adjustable interface, which is composed of multiple visualization layouts and several navigation controllers, to assist users' news search. Our tool provides users 1) multi-faceted network visualization, 2) recording or controlling of the browsing process, 3) related-term filtering, and 4) clues for the next step browsing.

2. Literature Reviews

2.1. Searching, classification and clustering

Users must form a suitable query when using search engines for their own purposes, so the process of information-seeking via keyword searching becomes very different from that of casual news reading or browsing which is relatively unintended. The study result of Delphi Group [7] indicated that for most users, search engine is an efficient way of retrieving network information, but more than half of them complained that the search results are too large and ill-organized for efficient retrieval. Moreover, many users spent more time browsing retrieval results rather than focusing on keyword retrieval.

The problem of reading massive search results can be solved by clustering or classification of news, both of which can provide users a map of framework to scaffold their news reading experiences. A useful classification scheme has at least two basic functions. First, with the help of clear hierarchy and connected concepts, users can easily observe the association among information, which serves as "external cognitive scaffolding [8]," and even form more new ideas [9]. The second function of classification is serendipitous guidance [10], which can help users discover relevant ideas that did not occur to them before. While classification defines sophisticated terms and meaningful relations among terms, clustering emphasizes more on less significant or invisible relationships such as "co-occurrence" or similarity of individual terms. A typical clustering process is to combine different sets or divide a set to increase the inner group similarity and the dissimilarity between groups [11].

Clustering mechanism has been applied to several information visualization tools such as *Vivisimo*, and *Grokker*. Käki and Aula [12] built up an information retrieval experiment to estimate the effect of clustering mechanism, and reported that those who use clustering mechanism can retrieve information faster than those who don't by 40%. They also noticed that users expect fewer clusters rather than many sophisticated clusters. And those who were familiar with clustering mechanism often use clustering results as queries later. However, most of these tools focus on building a better static map for visualizing information, instead of providing an adjustable and interactive interface for scaffolding the browsing process and needs.

2.2. Browsing strategies

The tool in our study aims to assist users in searching, browsing, and reading news. More specifically, we designed an interface to "navigate" users' news reading behaviors. Therefore, the tool should be able to implement some navigation functions in traditional user interfaces or web browsers. In Carledge and Pitkow's study [13] that investigated how users used the function of web browsers, they found in the system log that, the high frequency functions included (from high to low frequency) clicking hyperlinks, clicking the "backtrack" button, opening URLs, opening Hotlists, and forwarding. The frequency of clicking hyperlinks and the backtrack button was up to 93% in users' browsing behaviors. The most frequent browsing strategies are "Leave as you've entered" and "Looping back," both of which highlighted the importance of keeping historical clicking and browsing paths. While hyperlink, a directed link, leads users to jump from the current page to others even outside current sites, browsing path (or log), used to record the order of what a user has clicked, can avoid users losing in hyperlink jumping. However, since the browsing process is often a one-

way directed path, users can hardly jump back to the viewed pages. Therefore, all of the browsers provide browsing paths allowing users to track back to what they have read.

3. System Designs

To demonstrate the functions of our tool, *United Daily News* and *Chinatimes*, two Taiwanese online news sites, was crawled as the dataset, which included 793,522 documents from Jan 1st to Dec 31th in 2011.

3.1. Extracting key terms and building associations

To achieve this goal, we adopted Tseng, et al.'s method [6], which includes a series of process composed of stemming, term extracting, filtering, and association building. In this method, keywords were not extracted with language-dependent natural language processing, but a statistical algorithm, in which high frequency and successive words were merged together. The news documents were divided into two separate sets in advance to form stop word list and filter out the functional words. Regarding the process of building associations between two terms, Tseng, et al. suggested a new measure for connecting terms [6]. However, considering the purpose of navigation, we assumed that users may need to know the possible relations between one key term and others even with only one co-occurrence. Therefore, we allowed users to decide the threshold to broaden or narrow down the visibility of related terms.

3.2. Features of Interface

The main features of our tool are shown in Fig. 1. An online version is now available for testing (http://goo.gl/Jm26U).
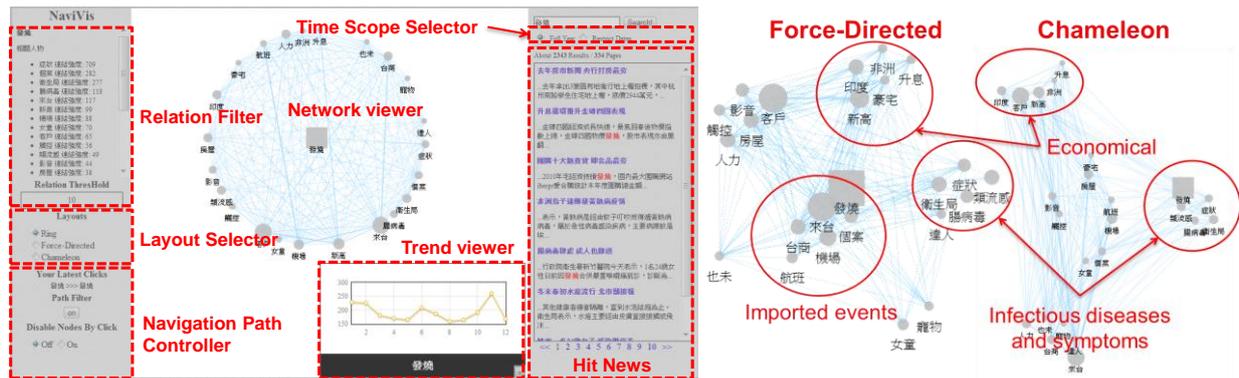

Fig. 1: An overview of the user interface. Right-sides are the other two graph drawing algorithm of network visualizers.

- *Network viewer.* Three network-based visualization layout algorithms were implemented, including ring, force-directed, and Chameleon layouts (see Fig. 1). In all three layouts, the area of each node (term) represents the amount of related hit news, while the thickness of a link (association) represents the amount of discovered associations. Ring layout provides users an overview of the related terms located with proper spacing on the edge of circle and the weight of associations between the query and its related terms, making it easy for users to find the strongly-related terms. Unlike the ring layout, force-directed layout locates each pair of terms with proper and reasonable distances but also avoids link crossing. The weight of associations between each pair of terms determines the location of the nodes and length of links between them. With the force-directed layout, users can not only discover the strong-related terms and highly-condensed subgroup of terms but also browse the next term more accurately and efficiently. The last algorithm, Chameleon [14], basing on agglomerative hierarchical clustering method, focuses on allocating clusters with proper distances between them, rather than allocating individual concepts directly as what force-directed layout does. In this case, users can see clearer clusters than those in the force-directed method. Finally, when using any of the three layouts, users can drag, click, and disable nodes.
- *Navigation path controller (click path, navigate forward or backward, clicked node set).* In our interface, users can interact with the visualized network by dragging or clicking nodes (terms). If a user clicks a related node t2 (for example, assuming t2 related to original t1), t2 will be expanded and searched by "OR" operator—that is, "t1 OR t2" as the next query after the user clicks t2. All of the

terms (whose weight of associations over the threshold) related to the 2nd query will be shown on layouts. To record users' browsing paths, the clicked terms and associations between them will be highlighted by color and shape and moved to the graph center (for example, square node in Fig. 1). With these featured designs, users can see what they clicked before, similar to the "historical log" function of Web Browsers such as Google Chrome. Besides visualizing clicking path (or set), the tool allows users to go backward by disabling the clicked nodes one by one. Once a clicked node is disabled, all of its related nodes, except for those also related to other clicked nodes, will be filtered out (Step II and III in Fig.2).

- *Filtering component.* We calculated the co-occurrence times of terms in sentences as weight of the association between each pair of nodes. The filtering model controls how many related terms and links will be displayed in the graph. Users can either broaden or narrow down the related terms with this option.

- *Time-related features viewer.* Users can use the period selector to focus on their interested events if they know when events occur. Moreover, the trend viewer right-below the network viewer displays the development of a certain event.

4. Case Study Results

We conducted a user study recruiting 10 university students in northern Taiwan to evaluate the usability, pros and cons of our tool. When the participants entered the system, they saw a brief introduction to the features and purposes of our tool. Immediately, they were led to a short personal background survey with items such as "Have you ever watched online news?" and "How often do you visit online news sites?" Then they were asked to browse news with our tool for at least 10 minutes. We used a free screen streaming software *CamStudio* to record their browsing process. Finally, they were interviewed and asked to share how they feel about our tool and what they discover through adjustable and interactive network layouts.

Fig. 2 shows the real browsing process of how a participant interacted with our interface. The participant accepted our suggestion and made the first query "flu" in Chinese. Then he clicked a related term of flu, that is, the name of the Director of Taiwanese Department of Health in 2011. However, he seemed uninterested in the term, so he disabled it. Afterwards, he clicked two terms in the successive two steps, filtered out the low-related terms, and finally changed the layout into *Chameleon*. After these operations, he started to read the related news in the right column as shown in Fig. 1. When we asked what he discovered, he said "I have never known that we are now suffering so many infectious diseases threatening our daily life."

According to the interview results, both the network layouts and navigation path controller received positive and negative feedback. Although many participants noticed that the clustering layout can summarize related terms and news, none of them mentioned how selectable layouts may help them read news in different views. Considering the interactive nodes of navigation path controller, some of the participants recognized disabling unwanted clicked nodes as a good feature, which provides them convenience to look back what they searched before. However, this feature to some extent confused parts of the participants. Besides, according to the recorded movie clips, none of the participant has used the time-range features, so they never discussed about it. One of the participants mentioned that the displaying logic of related terms in our tool is different from that of other search engines or user interfaces that were familiar to him.
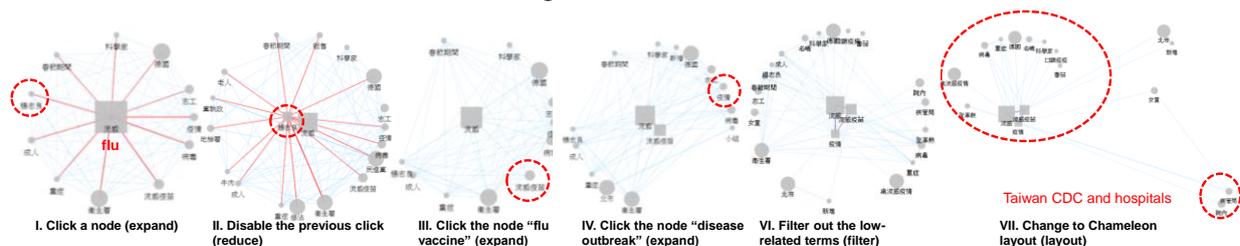

Fig. 2: A participant's successive six steps of operation.

5. Discussion and Conclusion

In conclusion, the computing power of current PCs and speed of network not only have raised the need of visualization tools but also benefited the development of visualization tools. When users consume large amount of data, visualization tools can lower down the threshold of information access and reduce the problem space in human daily information seeking and encountering. With the ultimate goal of navigating users to consume daily news in this paper, we designed news visualizers composed of several features such as interactive browsing paths and selectable network layouts, both of which provide users chances to encounter news unintentionally or find suitable related terms for their own purposes. The feedback of using the tool from our participants was both positive and negative. Most of them noticed that the designed features, especially the adjustable network layouts and interactive browsing path, can help them to encounter related news and improve the scope of related terms. Nevertheless, some of the participants also indicated that the tool contains too many functions that may confuse them, so several of them even preferred not to use such complex functions. While the current user study is preliminary, it is suggested that future research apply the ideas of our tool to specific domains, such as health news for normal people or technology news for students in college of science.

6. References

[1] D. Shenk. *Data smog: Surviving the information glut*. San Francisco: Harper Collins, 1997.

[2] M. J. Bates. The design of browsing and berrypicking techniques for the online search interface. Online Review. 1989, **13** (5): 407-431.

[3] H. H. Chen, J. J. Kuo, and S. J. Huang. A summarization system for Chinese news from multiple sources. *Journal of American Society for Information Science and Technology*. 2003, **54** (13): 1224-1236.

[4] S. Havre, E. Hetzler, and P. Whitney. Themeriver: Visualizing thematic changes in large document collections. IEEE *Transactions on Visualization and Computer Graphics.* 2002, **8** (1): 9-20.

[5] H. I. Chung, K. H. Liu, and T. Y. Li. Exploring dynamic social network through interactive visualization. *Proc. of* IEEE *Pacific Visualization*. 2010. Taipei, Taiwan.

[6] Y. H. Tseng, C. Y. Chang, and S. N. C. Rundgren. Mining concept maps from news stories for measuring civic scientific literacy in media. *Computers & Education*. 2010, **55** (1): 165-177.

[7] Delphi Group. *Information Intelligence: Content Classification and the Enterprise Taxonomy Practice*. 2004. Retrieved from http://www.delphigroup.com/whitepapers/pdf/20040601-taxonomy-WP.pdf

[8] E. K. Jacob. The everyday world of work: two approaches to the investigation of classification in context. *Journal of Documentation*. 2001, **57** (1): 76-99.

[9] Delphi Group. *Taxonomy and Content Classification*. 2002. Retrieved from http://www.delphigroup.com/research/whitepaper_request_download.htm.

[10] D. Bruno, and H. Richmond. The Truth about Taxonomies. *Information Management Journal*. 2003, **37** (2): 44-51.

[11] D. D. Lewis, and W. B. Croft. Term clustering of syntactic phrases. *Proc. of the 13th Annual International ACM SIGIR* Conference *on Research and Development in Information Retrieval*. 1990. pp. 385-404. Bruxelles, BE.

[12] M. Käki, and A. Aula. Findex: improving search result use through automatic filtering categories. *Interacting with* Computers. 2005, **17** (2): 187-206.

[13] L. Catledge, and J. Pitkow. Characterizing Browsing Strategies in the World-Wide Web. *Computer Networks and ISDN Systems*. 1995, **27** (6): 1065-1073.

[14] G. Karypis, E. H. Han, and V. Kumar. Chameleon: Hierarchical Clustering Using Dynamic Modeling. *Computer*. 1999, **32** (8): 68-75.