Exploring the World Wide Web with Self-Organizing Map

Christopher C. Yang
Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong, Shatin, Hong Kong
yang@se.cuhk.edu.hk

Hsinchun Chen
Department of Management Information Systems
The University of Arizona, Tucson, AZ 75721, USA

Kay Hong
Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong, Shatin, Hong Kong

ABSTRACT

Information overload is a critical problem in World Wide Web. Kohonen's self-organizing map (SOM) has been proven to be a promising browsing tool for the Web. The SOM algorithm automatically categorizes a large Internet information space into manageable sub-spaces. However, as the amount of information increases, it is expected to increase the size of the map accordingly in order to accommodate the important concepts in the information space. It results in increasing of visual load of the SOM. Large pool of information is packed closely together on a limited size of displaying window, where local details are difficult to be clearly seen. In this paper, we investigate the fisheye views and fractal views to support the visualization of SOM in order to explore the World Wide Web.

Keywords

Internet browsing, information visualization, fisheye view, fractal view, SOM, information overloading, visual load

1. INTRODUCTION

The Kohonen's self-organizing map (SOM) is a promising two-dimensional graphical tool for Internet browsing. Experiments showed that it is capable to categorize a large Internet information space, Yahoo Entertainment sub-category [**2**]. Given a large electronic information space, the SOM automatically categorize it into manageable sub-spaces so that users can navigate the map to locate the documents of interest. However, if the size of the SOM increases in order to fit the increasing numbers of categories, the details of the map become too small to be seen clearly. Effective visualization techniques are desired to solve such problem. The purpose of visualization is mapping information onto graphical representation to gain insight for users. Information visualization assists users to view and locate the information of interest in a limited space by interacting with the visualization system. Two visualization techniques, fisheye views and fractal views, are investigated in this paper. Fisheye views are developed based on the distortion approach while fractal views are developed based on the information reduction approach. The purpose of fisheye views are enlarging the regions of interest and diminishing the regions that are further away while maintaining the global structure. On the other hand, fractal views are approximation mechanisms to abstract complex objects and control the amount of information to be displayed. We have developed a prototype system and conducted a user evaluation to investigate the performance of fisheye views and fractal views. The results show that both fisheye views and fractal views significantly increase the effectiveness of visualizing SOM. In addition, fractal views are significant better than fisheye views but the combination of fractal views and fisheye views do not increase the performance comparing to each individual technique.

2. Browsing by SOM

Browsing is an exploratory and information-seeking strategy that depends upon serendipity, which is appropriate for ill-defined problems or exploring new task domains [14]. It is characterized by the absence of planning [13]. In essence, browsing explores both the organization and structures of the information space and its content, based on the pre-existing mental models of information organization. That's why it is often used in exploring relatively new or unexplored information spaces, such as World Wide Web. SOM clusters similar documents together and automatically assigns label to the categories. Similar documents, which are represented by similar noun phrase terms, are grouped together in a neighborhood. The SOM algorithm is summarized as follows:

1. Initialize input nodes, output nodes, and connection weights
2. Present each document in order
3. Compute distances between the input node and each output node

4. Updating weights of the winning output node and its neighbors to reduce their distance
5. Repeat Step 2 to Step 4 until converge
6. Label regions in map

An example of a 20 by 20 SOM is shown in Figure 1.

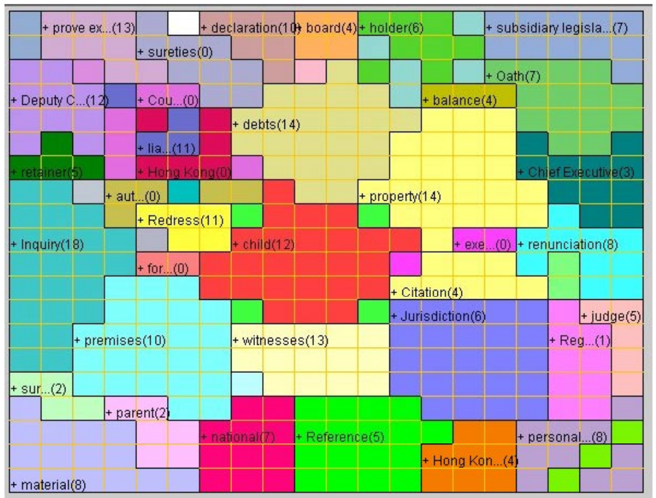

Figure 1. Sample of SOM.

3. Fisheye Views and Fractal Views

Fisheye views, first developed by Furnas [6] and further enhanced by Sarkar and Brown [1,15], are known as distortion techniques in information visualization [7]. Regions of interest are enlarged and the other regions are diminished so that one or more parts of a view are emphasized to show the importance of those regions. Both local details of the regions of interest and global structure of the overall display are maintained. By moving the point of interest, users may explore different areas of the two-dimensional map. When users are navigating the SOM, they may start from the concepts of interest and explore the similar concepts in the neighbourhood of the SOM. Fisheye views enable users to explore the region of interest and visualize the relationship of the further diminished regions. Figure 2 illustrates the effect of the fisheye view on the SOM as shown in Figure 1.

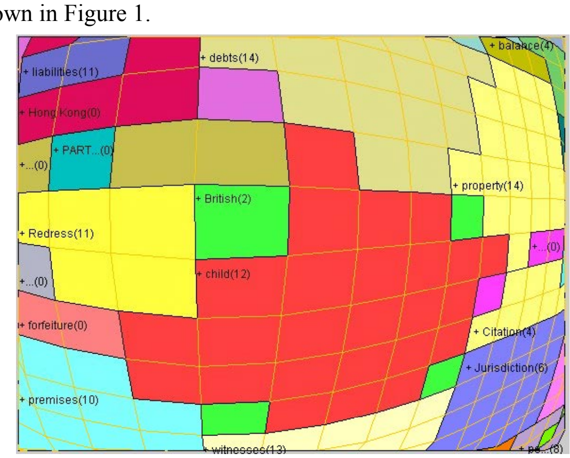

Figure 2. Fisheye View

Fractal views are information control methods for information display. It uses the information reduction approach to control the amount of information displayed by focusing on the syntactic structure of the information. Fractal is an important concept in fractal theory [12] to describe the complexity of an entity in both quantitative terms and mathematical terms. Fractal view [5] is an approximation mechanism to abstract complex objects and control the amount of information to be displayed with a scale (threshold) set by users. Using such mechanism, the total amount of information displayed is consistent given any focuses of attention but only details near the focus point and important landmarks further away are displayed. The amount of information displayed is also flexible to the interest of users. Figure 3 illustrates the effect of fractal view on the SOM as shown in Figure 1.

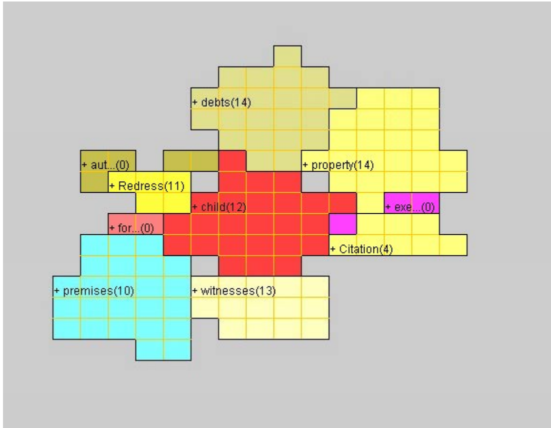

Figure 3. Fractal View

4. Experiments

An experiment has been conducted to compare the effectiveness of the visualization techniques, fisheye views and fractal views, on SOM. This study analyzed how different visualization techniques affect the performance of information searching tasks on SOM. Visualization techniques are evaluated with respect to their suitability for certain task characteristics [8,9]. Task-based experiments are commonly used in comparing the performances of different visualization techniques [16].

Twenty subjects were recruited to participate in the experiment. No subjects have any prior experience of SOM browser and visualization tools. However, a training session was given to all subjects before they conduct the test. There were totally four setups of the SOM visualization techniques, 1) fisheye views, 2) fractal views, 3) combination of fisheye views and fractal views, and 4) no visualization technique. Each Subject was assigned to use the four setups in a random order. The Hong Kong government Web site is selected as the Web information space to be explored in the experiment. A 20 by 20 SOM is automatically generated to categorize the Web pages of this site. Three interface panels, 1) SOM interface, 2) control interface, and 3) documents interface, are available in the experiment prototype system as shown in Figure 4. When users click on a region in the SOM interface panel, the documents of the region will be presented in the document interface panel together with the documents' keywords and titles

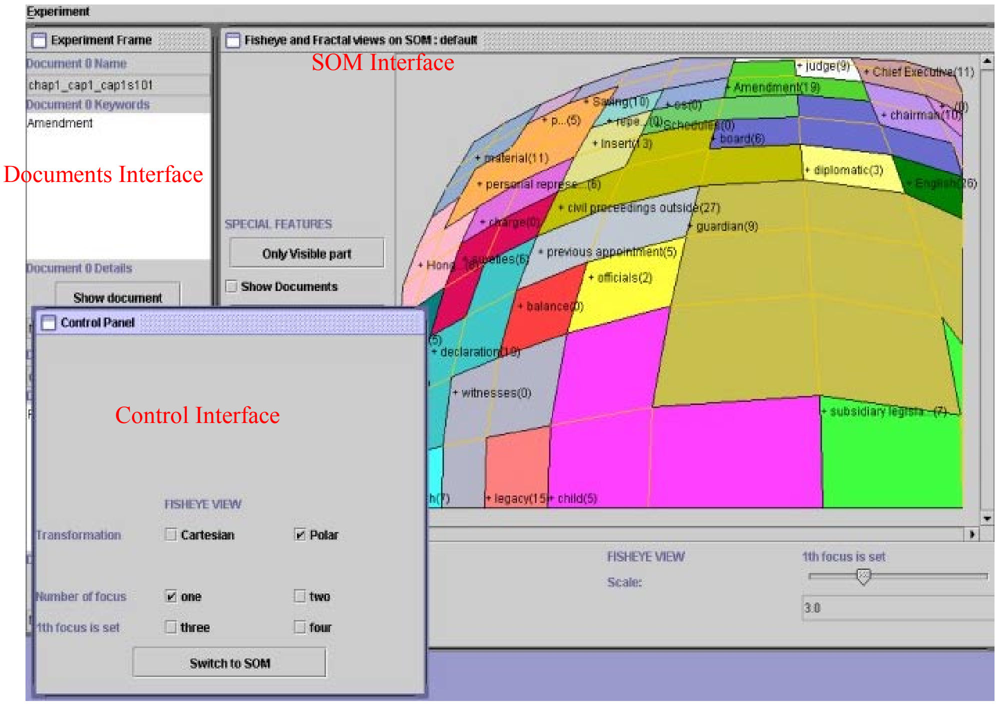

Figure 4. System interface of the experiment.

4.1 Results and Discussion

The average time taken for twenty subjects to complete the assigned tasks using four different setups is shown in Figure 5. A one-way analysis of variance (ANOVA) is conducted to analyze if there is any significant difference between the performances of different visualization techniques. The result is shown in Table 1. The p-value is 0.018. That means the difference between the time taken to complete tasks by different visualization technique is significant at the level of 0.018.

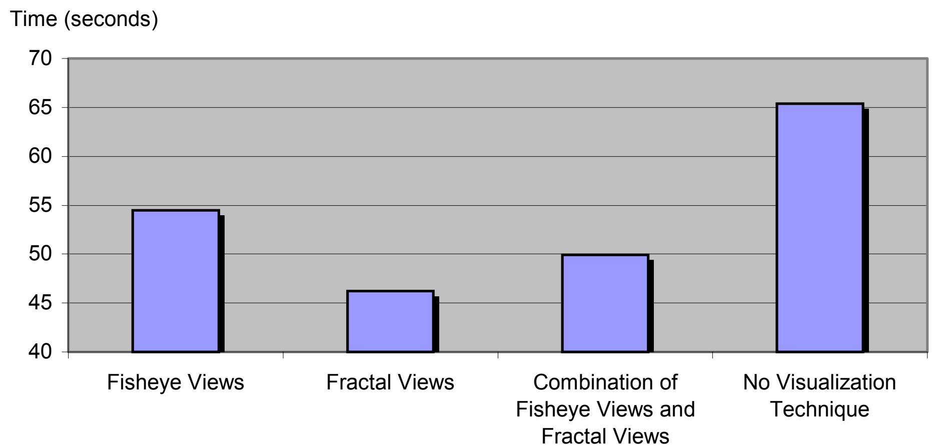

Figure 5. Experimental results.

Table 1 ANOVA result of the time taken between different visualization techniques

	Sum of Squares (SS)	Degree of Freedom (df)	Mean Sum of Square (MSS)	F-ratio (F)	p-value (p)
Between different visualization techniques	2550.508	3	850.1693	3.583764	0.017567
Within different visualization techniques	18029.33	76	237.228		

Since there is significant difference between using different visualization techniques, we conduct the t-test to investigate the difference among the visualization techniques. The results are shown in Table 2 and Table 3. We find that the time taken to accomplish the tasks by using fractal views is significantly less than that using no visualization technique ($p = 0.006728$). However, the difference between using fisheye views and no visualization techniques is comparatively less significant ($p > 0.05$). We also find that the difference between using fisheye views and using fractal views is significant at the level of 0.0787, but there are no significant difference between using fisheye views and combination of both and between fractal views and combination of both.

Based on the above observations, the information reduction approach (fractal views) is more effective that the distortion approach (fisheye views) in visualization of SOM although both are useful. Maintaining the global structure is important in visualization, but reducing the visual load is more helpful to users in order to explore the areas of interest. The major drawback of fisheye views is the distortion of the original shape of regions in SOM. On the other hand, fractal views require less mental integration between the views because there are only additions or reductions of regions being displayed but shapes of regions are not changed.

5. Conclusion

SOM has been proven a powerful browsing tool for large information space and World Wide Web. However, visualization of SOM becomes less effective as the size of the map increases to accommodate more concepts of the information space. In this paper, fisheye views and fractal views are investigated. Fisheye views distort the map to present the local details and maintain the global structure. Fractal views reduce the information presented on the map by keeping the most relevant regions of the region of interest. A user evaluation has been conducted to investigate the performance of the proposed techniques. The experiment results show that both fisheye views and fractal views improve the effectiveness of visualization significantly. However, combining both techniques do not produce any significant difference comparing with the individual techniques. In addition, fractal views perform significantly better than fisheye views.

Table 2 **T-test between fisheye views and no visualization technique, between fractal views and no visualization technique, and between combination of both and no visualization technique**

	p-value
Between fisheye views and no visualization techniques	0.057383
Between fractal views and no visualization techniques	0.006728
Between combination of fisheye views and fractal views and no visualization techniques	0.009373

Table 3. **T-test between fisheye views and fractal views, between fisheye views and combination of both, and between fractal views and combination of both**

	p-value
Between fisheye views and fractal views	0.077897
Between fisheye views and combination of fisheye views and fractal views	0.202655
Between fractal views and combination of fisheye views and fractal views	0.254825

6. Acknowledgement

This project is supported by the Earmarked Grant for Research of the Research Grants Council of Hong Kong, CUHK 7034-98E.

References

1. M. H. Brown, J. R. Meehan, and M. Sarkar, “Browsing Graphs Using a Fisheye View,” *Proceeding of ACM on Human Factors in Computing Systems*, Amsterdam, Netherlands, (1993).
2. H. Chen, A. L. Houston, R. R. Sewell, and B. R. Schatz, “Internet Browsing and Searching: User Evaluations of Category Map and Concept Space Techniques,” *Journal of the American Society for Information Science*, vol.49, no.7, (1998), pp.582-603.
3. H. Chen, C. Shuffels, and R. Orwig, “Internet Categorization and Search: A Machine Learning Approach,” *Journal of Visual Communications and Image Representation*, vol.7, no.1, (1996), pp.88-102.
4. H. Chen, and K. J. Lynch, “Automatic Construction of Networks of Concepts Characterizing Document Databases,” *IEEE Transactions on Systems, Man, and Cybernetics*, 22, (1992), pp.885-902.
5. J. Feder, *Fractals*, Plenum, New York, (1988).
6. G. W. Furnas, “Generalized Fisheye Views,” *Proceedings of the SIGCHI Conference on Human Factors in Computing System*, (1986).
7. M. Heo, and S. C. Hirtle, “An Empirical Comparison of Visualization Tools to Assist Information Retrieval on the Web,” *Journal of the American Society for Information Science and Technology*, vol.52, no.8, (2001), pp.666-675.
8. D. A. Keim, “An Introduction to Information Visualization Techniques for Exploring Very Large Databases,” *Tutorial Notes*, Information Visualization '00, Salt Lake City, Utah, USA, October, (2000).
9. D. A. Keim, “Visual Exploration of Large Date Sets,” *Communications of The ACM*, Vol.44, No.8, August, (2001), pp.39-44.
10. T. Kohonen, “The Self-Organizing Map,” *Proceedings of the IEEE*, vol.78, no.9, (1990), pp.1464-1480.
11. H. Koike, “Fractal Views: A Fractal-Based Method for Controlling Information Display,” *ACM Transactions on Information Systems*, vol.13, no.3, (1995).
12. B. B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman and Company, New York, (1977).
13. G. Marchionini, “An Invitation to Browse: Designing Full Tex System for Novice Users,” *Canadian Journal of Information Science*, vol. 12, no.3, (1987), pp.69-79.
14. G. Marchionini, and B. Shneiderman, “Finding Facts vs. Browsing Knowledge in Hypertext Systems,” *IEEE Computer*, vol.21, no.3, (1988), pp.70-79.
15. M. Sakar, and M. H. Brown, “Graphical Fisheye Views of Graphs,” *Proceedings of ACM on Human Factors in Computing Systems*, Monterey, CA, (1992).
16. D. Turo, and B. Johnson, “Improving the Visualization of Hierarchies with Treemaps: Design Issues and Experimentation,” *Proceedings of the IEEE International Conference on Visualization*, Los Alamitos, CA, USA, (1992), pp. 124-31.
17. C. C. Yang, J. Yen, and H. Chen, “Intelligent Internet searching agent based on hybrid simulated annealing,” *Decision Support Systems*, Vol.28, Issue 3, May, (2000), pp.269-277.