

# Client-Side Visualization of Internet Forums for Information Retrieval

Guangfeng Song

Penn State Greate Allegheny, 4000 University Dr, McKeesport PA 15132, USA  
gus4@psu.edu

**Abstract.** This paper presented a method to visualize Internet forums on the client side. Enhancement to web browsing was proposed to solve problems in information retrieval from forums. Specifically, visualization of structural information provided useful overviews for web browsing. A conceptual model described forums as three-dimensional spaces with information objects as points and web browsing as movements in the spaces. A series of diagrams were proposed to provide overviews for users' web browsing movements. Implementation of the visualization system and examples of the diagrams were presented in the paper.

**Keywords:** Visualization, Internet Forums.

## 1 Introduction

Internet forums are websites designed to hold user-generated. Each web-based forum becomes a virtual community of users with shared interest in a certain area. Anonymous viewing of a forum is possible but users typically need to register as members before making posts. An initial post on a topic, along with all the posts replying the initial post, becomes a thread. These threads are organized into forums by their topics.

Internet forums are web sites where users are content creators as well as content consumers. There is an abundance of information on almost any topic. Sometime useful but unofficial information can only be found in forums. While most forums contain personal opinions with varied degree of creditability, some forums have more authoritative content. For example, the Linux distribution Ubuntu relies on its forums to provide user support. Forums are useful to find a solution to a problem that other people may have experienced, as more and more people are contributing their solutions on the Web. The usefulness of forums grows with the abundance of information. However, the actual use of forums requires successful retrieval of specific pieces of information, which in turns depends on two tasks: locating an appropriate forum and locating appropriate posts in the forum. Search engines can help users complete these two tasks in one step, provided that the correct search terms are supplied. Otherwise, users will have to complete the tasks with web browsing, which can be slow and demanding information processing tasks. The problems in retrieval of information from forums can be summarized as the followings:

1. Lack of effective quality control in forums
2. Difficult to search information in forums by timestamp
3. Rigidity of topics organization in forums
4. Inability to search for communication and collaboration
5. Deficiency of search engines in searching forums

Much of the above problems can be blamed on the inadequate utilization of attributes of the forum objects. A rigid organization by topics is often the only visible forum structure to users. While other attributes such as authors, timestamps, and size of posts are displayed along the posts, these attributes are not used to organize forums. Nor can they be used to quickly browse the forums. Search engines are limited because they do not utilize these attributes either. Moreover, search results as long list of web pages destroys the familiar look of forums. Navigation of the long list of search result can be slower than following the structure of forums.

The alternative as proposed in this paper is to enhance the usability of browsing forums by utilizing all attributes of forum objects. A well-known weakness of web browsing is the loss of location awareness which is commonly referred as lost in cyberspace [1]. This occurs because users do not have an overview of cyberspace which is hard to generate for websites of pages without well-defined structures. A generic network graph with identical nodes of web pages will not be a useful overview for users. However, Internet forums do have structures that enable more useful overviews. The spatial relationships among entities in a graphical overview will be easy for human to understand. More broadly, many information visualization techniques can be used to enhance the ability of users to browse forums.

## 2 Information Visualization

Shneiderman provided a taxonomy of visualization tasks that has been widely adopted [2]. Visualization tasks can be classified into the following aspects of information presentation: overview of information, zooming, filtering, details on demand, relationship among items, history of actions, and extraction of subset.

Despite the theoretic benefit of information visualization, experimental studies have been mixed on the benefit of visualization in reducing task completion time [3]. This discrepancy is likely due to complex interaction between task factors, user characteristics, and visualization techniques. A mismatch of these items may be the cause of the problems [4].

Therefore, successful visualization depends on appropriate use of visual primitives to represent information. These primitives include: length, area size, volume, relative location, shape, density, color, and others. Intrinsic meanings have emerged for some attributes have developed due to their consistent use in people's real life experience. For example, larger size (length, area, or volume) represents larger value, darker density or color intensity refers to larger values, leftmost position corresponds to first item, and rightmost position corresponds to last item [5]. Colors are most often used in visualization to distinguish one element from another [6]. There's no intrinsic meaning for color hues but the difference of colors can be easily perceived. Hence, contrast of colors is useful to distinguish groups while similar colors aid grouping.

However, overuse of colors for grouping can lead to clutters and should be avoided [6]. A classification of many visual primitives can be found in [7].

There are a large variety of metaphors to present data with selected visual primitives. They range from simple diagrams such as simple bar charts, to hierarchical bar charts [8], and to metaphors of real world such as People Garden [9]. A classification of simple visualization metaphors can be found in [10]. Successful metaphors require good mapping between data attributes and the visual primitives in a setup that easily perceivable by users. Constant exploration of new metaphors is part of progress researchers are making in the field of information visualization.

Visualization of Internet forums is relevant to visualization of social networks. Freeman reviewed the roles of visualization in analyzing social networks [11]. Researchers often concentrated on the network structure emerged from specific user communities of collaboration groups [9][12]. Visualization studies have also explored the temporal dynamics of communication flows in social networks. In [12], Flows of communication were generated from a mailing-list by mining email log files. Graphs over a period of time are merged into movies to highlight the temporal aspects. The visualization is similar to the weather map in weather forecast. The emphasis is on the progressive changes in shape and connectivity of graphs. The use of animation is common practice in visualization of temporal attributes [10].

Multi-dimensional information is a natural result of using multiple attributes of objects in visualization. This is exactly the case with Internet forums. However, high dimensionality must be reduced before visualization is possible. The dimensions are usually reduced to two or three by multidimensional scaling methods. Dwyer and Gallagher visualized a high dimensional portfolio holding data set with attributes that change over time. Their visualization was achieved with dimension reduction [13]. Other studies in visualization of multidimensional data include "Table Lens" for data in a table [14] and WebTOC [15] for multiple attributes of all web pages in a website.

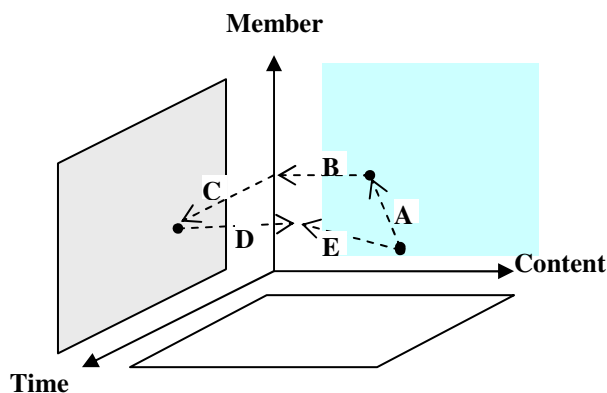
### 3 Methods to Visualize Internet Forums

The model of internet forums is a mix of hierarchy and linear arrangement. The basic objects in a forum are posts made by members. Posts of a topic are arranged by time in a linear sequence. The topics or threads are also arranged by time in a linear sequence and they are placed in a branch of the hierarchy of forums. Therefore, Internet forums have the hierarchy of posts-threads-forums and the linear sequence of posts and threads.

The information in a forum has many natural attributes. A post has its author, post time, length, and other attributes. In addition, many aspects are often necessary to describe the text content of a post. Hence, the interface of a forum has the difficult task of presenting complex information of multiple attributes to users. In particular, the interface has to effectively convey these attributes under the limitation of 2-d display. Unfortunately, the roles of these attributes in facilitating information retrieval from Internet forums are often ignored in current forum interfaces.

Conceptually, members, content (members' posts), and time are three major "dimensions" of Internet forums. With this notion of three dimensions, each forum post becomes a point in a three-dimensional space. Browsing a forum is therefore modeled

as a sequence of movements inside the three-dimensional space. A move from one point to another in the space can be one of the three types: moves along one dimension while keeping other dimensions constant (Type I), moves within a two-dimensional plane while keeping the third dimension constant (Type II), and moves that involve changes along all three dimensions (Type III). In Figure 1, B,C, and D are Type I moves while A is a Type II move. Type III moves can be simulated by combining Type I and Type II moves such as E in Fig. 1.



**Fig. 1.** Dimensions in Internet Forums

A list of visualization methods are presented in Table 1. Each method is a visualization of either a Type I move or a Type II move. Type III moves are not directly visualized, however, because they would require 3-d diagrams. We intentionally limit the choice of diagrams to two-dimensional for several reasons: Computer interfaces are inherently two-dimensional; the value of 3-d visualization has not been convincing [16][17]; and Type III moves can be effectively simulated by these 2-d diagrams. These interfaces serve to uncover relationships among dimensions that help the search for relevant information by users.

## 4 Implementation of a Client-Side Visualization System

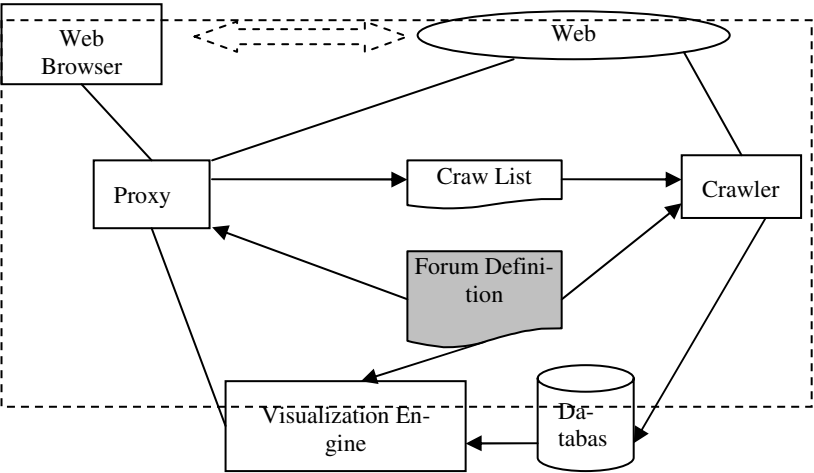
Fig. 2 illustrates a client-side visualization system for Internet forums. The system comprised of a proxy server that serves three purposes when users are browsing a forum: relays information from the Web, displays available local visualization of the forum, and instructs the crawler to retrieve more information from the forum. The crawler receives definition of the forums and a list of web addresses to process. A forum definition file enables the crawler to understand the content and structure of web pages in the forum so that structured data can be retrieved into a database. The visualization engine create diagrams discussed in previous section as requested, using the database and the forum definition.

**Table 1.** Visualization of Type I and II Browsing Moves

Dimension 1	Dimension 2	Specific Information	Visualization
Content	Content	Hierarchy of forums	V0
		Threads by threads similarity	V1
		Posts by posts similarity	V2
	Time	Threads by Time	V3
		Posts by Time	V4
	Member	Similarity of posts by members	V5
		Threads started by members	V6
		Participation of members in threads	V7
Member	Member	Members to Members quote	V8
	Time	Members' participation by time	V9

**Table 2.** Diagrams to implement the visualization for browsing moves

Visualization	Current Interface	Visualization Diagram
V0	Text links	Treemap of forums and sub-forums
V1	N/A	2-D Scatter plot with threads as nodes
V2	N/A	2-D Scatter plot with posts as nodes
V3, V4	Text lines	Multiple lines of bar charts along timeline (sparklines [])
V5	N/A	2-D Scatter plot with posts as nodes
V6, V7	Text pages	Correspondence diagram between nodes of members and threads
V8	N/A	2-D Scatter plot with members as nodes
V9	N/A	Multiple lines of bar charts along timeline



**Fig. 2.** Implementation of the visualization system on the client-side

The proposed system was tested on a popular forum avsforum.com. A number of visualization diagrams were generated. Examples are shown below.

**Hierarchy of forums.** This is a Treemap diagram that describes forums and sub-forums (V0 in Table 2). Categories of forums are distinguished by different colors. Each rectangle is a forum and the area size of the rectangle represents number of threads in the forum. Tooltips of forum information are displayed while the mouse is hovering over rectangles. A click of a rectangle bring users to list of threads along timeline (Fig. 3).

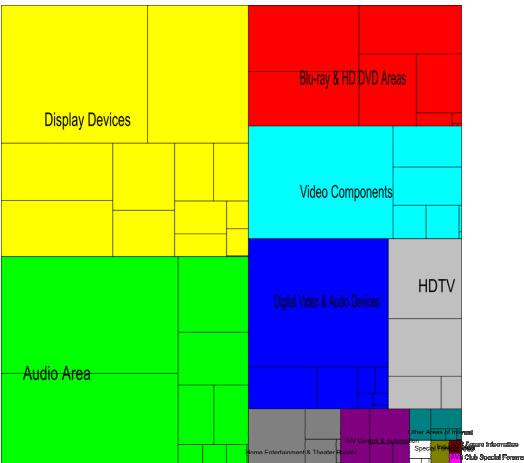


Fig. 3. Treemap diagram for hierarchy of forums

**Similarity diagrams.** These are 2-D scatter plots as shown in Fig. 4. This diagram is similar in appearance to "Topic Islands" in [18]. Each circle represents a post. The size of circles corresponds to the size of posts. Closeness of the circles corresponds to similarity of among the posts. Mouse hovering over a circle reveals information of the corresponding post. Clicking a circle bring users to the web page of the post. The similarity diagrams may be used to implement V1,V2,V5 in Table 2, when colors are used to represent different meanings.

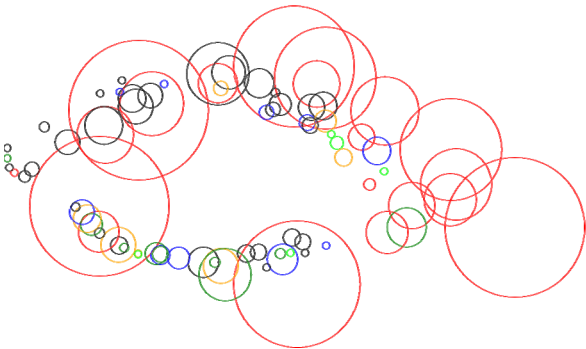


Fig. 4. Scatter plot for similarity among posts

**Timeline diagrams.** Two types of timeline diagrams were created: one for visualizing multiple threads along the timeline (V3 in Table 2, shown in Figure 5) and one for visualizing posts inside a thread (V4 in Table 2, shown in Figure 6). In both diagrams, the height of bars represents the size of the posts and the location of the bars represents the time when the posts were created. Colors were used to distinguish among members who made most posts within a thread. The timeline diagrams was also used to implement the posts made by a member over time (V9 in Table 2) when color represented different members. These diagrams are similar to Sparklines in providing overviews [19].

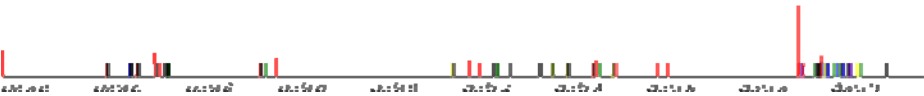


Fig. 5. Timeline diagram for posts in a thread

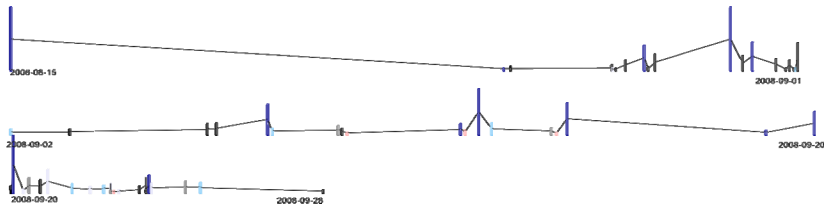


Fig. 6. Connected timeline diagram for posts in a thread by pages

## 5 Limitations and Conclusion

The main limitation of the proposed visualization approach is the forum definition of semantics that must be performed manually. Although software tools have been developed to ease the process, it is still a limiting factor to visualize large number of Internet forums and changes in the forums may require new definition files. Another limitation is the potential large amount of data that has to be downloaded and stored locally for the visualization to work. Therefore, the visualization method in this paper may be most appropriate for users who only browse a few forums frequently. On the other hand, server-based visualization solves these problems nicely as the data and data structure is readily available on servers. But server-based visualization is less flexible and may significantly increase the burden on servers because all visualization has to be generated by servers. Both client and server-based approaches shall be pursued in future work.

In conclusion, this paper presents a client-side visualization solution to the problem of information retrieval from Internet forums. Forums are modeled as three-dimensional space and browsing forums is equivalent of moving in the space. The proposed visualizations enhance the ability of users to quickly browse for information in forums. The client-side implementation is a highly flexible system that can easily support any forums with any number of visualizations.

## References

1. Brake, D.: Lost in cyberspace. *New Scientist* 154(2088), 12–13 (1997)
2. Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: *Proceedings of IEEE Symposium on Visual Languages*, pp. 336–343 (1996)
3. Zhu, Y.: Measuring Effective Data Visualization. In: *Advances in Visual Computing*, pp. 652–661. Springer, Berlin (2007)
4. Sebrechts, M.M., Cugini, J.V., Laskowski, S.J., Vasilakis, J., Miller, M.S.: Visualization of search results: a comparative evaluation of text, 2D, and 3D interfaces. In: *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, Berkeley, California, United States, pp. 3–10 (1999)
5. Senay, H., Ignatius, E.: Rules and principles of scientific data visualization. *State of the art in data visualization (ACM) SIGGRAPH Course Notes* (27) (1990)
6. Healey, C.G.: Choosing effective colours for data visualization. In: *Proceedings of the 7th conference on Visualization 1996*, p. 263. IEEE Computer Society Press, San Francisco (1996)
7. Wünsche, B.: A survey, classification and analysis of perceptual concepts and their application for the effective visualisation of complex information. In: *Proceedings of the 2004 Australasian symposium on Information Visualisation*, vol. 35, pp. 17–24 (2004)
8. Keim, D.A., Hao, M.C., Dayal, U.: Hierarchical Pixel Bar Charts. *IEEE Transactions on Visualization and Computer Graphics* 8(3), 255–269 (2002)
9. Xiong, R., Donath, J.: PeopleGarden: creating data portraits for users. In: *Proceedings of the 12th annual ACM symposium on User interface software and technology*, Asheville, North Carolina, United States, pp. 37–44 (1999)
10. Senay, H., Ignatius, E.: A knowledge-based system for visualization design. *Computer Graphics and Applications*, IEEE 14(6), 36–47 (1994)
11. Freeman, L.C.: Visualizing Social Networks. *Journal of Social Structure* 1(1), 4 (2000)
12. Gloor, P.A., Zhao, Y.: TeCFlow-A Temporal Communication Flow Visualizer for Social Networks Analysis. In: *CSCW 2004 Workshop on Social Networks* (2004)
13. Dwyer, T., Gallagher, D.R.: Visualising changes in fund manager holdings in two and a half-dimensions. *Information Visualization* 3(4), 227–244 (2004)
14. Rao, R., Card, S.K.: The table lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. In: *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence*, Boston, Massachusetts, United States, pp. 318–322 (1994)
15. Nation, D.A., Plaisant, C., Marchionini, G., Komlodi, A.: Visualizing websites using a hierarchical table of contents browser: WebTOC. In: *Proceedings of the 3rd Conference on Human Factors and the Web* (1997)
16. Ridsen, K., Czerwinski, M.P., Munzner, T., Cook, D.B.: An initial examination of ease of use for 2D and 3D information visualizations of Web content. *Int. J. Hum.-Comput. Stud.* 53(5), 695–714 (2000)
17. Cockburn, A., McKenzie, B.: 3D or not 3D?: evaluating the effect of the third dimension in a document management system. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 434–441. ACM Press, New York (2001)
18. Miller, N.E., Wong, P.C., Brewster, M., Foote, H.: TOPIC ISLANDS - A Wavelet-Based Text Visualization System. In: Ebert, D., Hagen, H., Rushmeier, H. (eds.) *IEEE Visualization 1998*, pp. 189–196 (1998)
19. Tufte, E.R.: *Beautiful Evidence*. Graphics Press (2006)