Ripple Presentation for Tree Structures with Historical Information

Masaki Ishihara **Kazuo Misue** **Jiro Tanaka**

Department of Computer Science,
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki, 305-8573, Japan

{ishihara, misue, jiro}@iplab.cs.tsukuba.ac.jp

Abstract

We propose a new method for representing tree structures with historical information. We call this method Ripple Presentation. Categories of nodes are represented by the angles of edges and elapsed time is represented by the length of the edges. In this way, the method can express both the time series and categories, which has been difficult to achieve with either tree structures or lists. As a result, users can understand the overall information from their viewpoint view and discover target information effectively. We applied the method to trackback links of Weblog articles and the latest articles of News sites using RSS on Web as a resource.

Keywords: Information Visualization, Graph Drawing, Historical information, Tree structures.

1 Introduction

Finding information on the Internet can be a very time-consuming task. Even if we can obtain a list of relevant Websites using a search engine, the latest information is often difficult to find. This is particularly problematic when there is an urgent need for the latest information.

Recently, an increasing number of Websites distribute their content using RSS (Rich Site Summary). RSS provides indexes and summaries of Websites and updated information. Updated information from multiple Websites can be easily obtained using tools such as RSS Reader and News Reader.

However, information is often displayed using a tree or list in a time-series order, so finding the exact information from the perspective of category and time still takes time. We believe that there is need for and improved and more intuitive approach to finding relevant and up-to-date information.

In this paper, we address the problem of effective discovery of requested information by introducing approach of visualizing a large amount of information on the Internet.

The tree structure is widely used as a method of classifying information. It is commonly used for computer directories, the class structure of JAVA, organization charts, and distribution diagrams. The advantage of the tree structure is its ability to organize information in categories.

Once the root is decided, the tree structure becomes a directed graph. For instance, each hierarchy shows the unit of organization in organization charts or the age in distribution diagrams.

The weakness of the tree structure is that the latest information is always laid out at the bottom of the hierarchy because of its drawing convention when the hierarchy means time-series. Furthermore, the link relations between leaf nodes are not fundamentally expressible due to a tree structure that does not have closed loops.

In a tree structure, information can also be sorted by time. However, it then becomes difficult to organize the information in categories.

We propose an expression method called Ripple Presentation. As the name suggests, as time passes, older information (historical information) is moved outwardly away from a central point, where new information is positioned first. Though the targeted data structure is the tree structure, our method differs from the past method in two points:

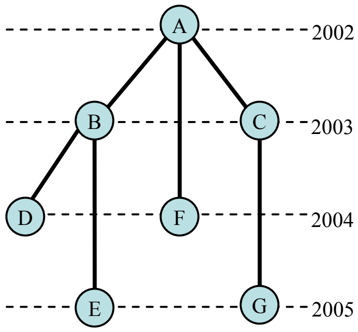
1. The latest information is positioned in the neighborhood of parent nodes.
2. The relationships between child nodes are indicated by the angles of edges from the parent nodes.

2 Outline of Ripple Presentation

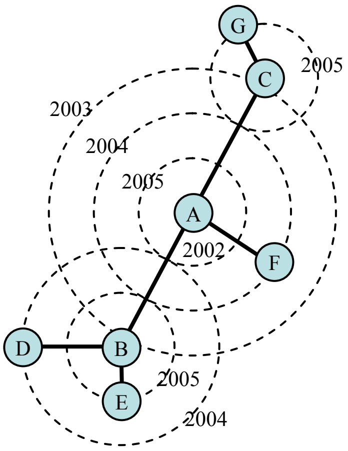
The concept of Ripple Presentation is that older information is positioned far away from and newer information closer to the user's viewpoint.

In the expression of a tree structure with a time-series hierarchy, such as in a distribution diagram, the hierarchy of the time series is normally defined in a parallel manner from the root node to leaf nodes as shown in Fig. 1(a). In Ripple Presentation, concentric circles (ripples) whose centers are the parent nodes of each subtree define the hierarchy as shown in Fig. 1(b).

Furthermore, the arrangement of the hierarchy is different in the two methods. In a tree structure with the time series, the hierarchy is generally arranged from the root node toward leaf nodes in ascending order. In our approach, it is arranged from leaf nodes toward the parent nodes of each subtree in the same order. Only the position of the root nodes is decided at its creation time. In other words, the arrangement of the time-series hierarchy is opposite to the arrangement of a tree structure with a time series.



(a) Time-series Hierarchy Structure of Tree



(b) Ripple Presentation

Fig. 1: Comparison of Time-series Hierarchy Structure with Ripple Presentation for Tree.

Another analogy in the natural world is the annual rings of a tree. An annual ring represents one year of the growth cycle, and the interval between the ripples is proportional to the reciprocal of the frequency of the wave. Although the cycles vary, the common point is that the interval has a constant cycle (assuming that there is no attenuation) and its shape is a concentric circle.

In Ripple Presentation, the cycle of each ripple is constant. Assuming that the attribute time of the node is the same as the creation time of the ripple, the node is positioned on the spreading ripple because the ripples are synchronized with the nodes. The parent nodes are positioned at the center of the ripples, and the ripples are generated at certain constant intervals as concentric circles. That is, all nodes on the same ripple means the nodes were generated at the same time.

Thus, once the time cycle of each ripple is understood, the history of the tree from some point in time to the oldest time is understood. In addition, it is only necessary to pay attention to the neighborhood of the generation points of the ripples (parent nodes) in order to know the latest information (leaf nodes).

3 Layout Design

This section explains how to lay out the tree structure data by Ripple Presentation in detail. First the targeted data is modeled and the layout design is mathematically explained using a graph model.

3.1 Rooted Tree with Historical Information

Our system targets a rooted tree with historical information. Suppose that $G = (V, E)$ is a rooted tree. Sets $E \subset V \times V$ of edges and sets V of nodes compose G. Sub trees described with G_{sub} are rooted trees composed of only the leaf nodes as the child nodes. Each node $v \in V$ of G includes some attributes such as category and creation time in the system. In this case, V represents the time series data. The creation time of v is described as $d(v)$.

3.2 Drawing Conventions

According to the strict definition (K. Sugiyama 2002) drawing conventions are composed of placement conventions for nodes and routing conventions for edges.

3.2.1 Placement Conventions for Nodes

In Ripple Presentation, the placement convention for G_{sub} is a *concentric circle*. The placement convention for G is a *recursive concentric circle*, since G is made up of G_{sub} as a nested structure.

3.2.2 Routing Conventions for Edges

The routing conventions are explained by line type and the relationship with the coordinate system. In Ripple Presentation, the line type is *straight line routing* and routing is *independent of coordinate system*.

3.3 Drawing Ripples

Ripples start generating when child nodes are created. That is, assuming that the oldest node (the most outside node) in all nodes except for root node that composes a subtree is v, ripples start generating at time $t_0 = d(v)$.

The ripples are not only generated synchronously with child nodes but also at a constant frequency. Therefore, by viewing the ripples, users can be aware of the elapsed time to accommodate the length of the edges. More accurately, from the projection lines (time scale) provided by the ripples, users can understand visually the scale of the elapsed time.

Suppose that $\{c_1, c_2, \cdots, c_n\}$ is a set of ripples centered on a parent node of subtree G_{sub}. At time t, the radius $r_1(t)$ of ripple c_1 (the most outside ripple) generated at time t_0 is expressed by

$$r_1(t) = \lambda \cdot f \times (t - t_0) \quad (3.1)$$

where f is the frequency of the ripples and λ (common parameter in all the ripples) is their wavelength. Then, the number of ripples n is

$$n = \left\lceil \frac{r_1(t)}{\lambda} \right\rceil \quad (3.2)$$

Therefore, at time t, the radius $r_i(t)$ of ripple c_i (inside ripples from c_1) generated after time t_0 is defined as

$$r_i(t) = r_1(t) - \lambda(i-1) \quad (i = 1, 2, \cdots, n) \quad (3.3)$$

3.4 Layout of Nodes

A polar coordinate system is used for Ripple Presentation as shown in Fig. 2. To put it more concretely, the length and the angle of the edges are required as parameters at the time of drawing G, and then, based on these parameters, the position of each node is decided. For example, position $p(v)$ of child node v is defined by polar representation $(l_v(t), \theta)$, where the length of edge $e \in E$ is $l_v(t)$ and the angle is θ. That is, positions of all nodes are relatively decided when the position of the root node is decided.

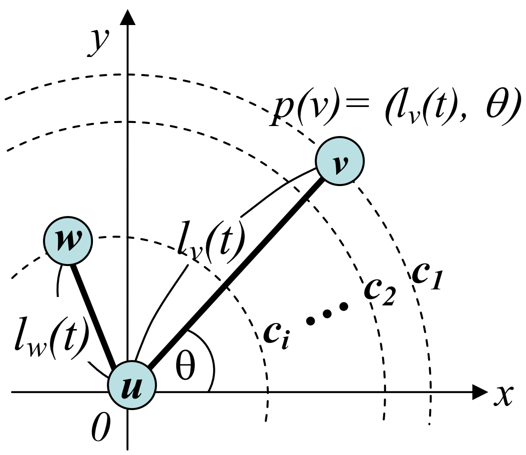


Fig. 2: Polar Representation of Child Nodes.

3.5 Length of Edges

At time t, the length $l_w(t)$ of the edge from parent node $u \in V$ to child node $w \in V$ is defined as

$$l_w(t) = r_1(t) - \lambda(j-1) \quad (j = \lceil f \times (d(w) - t_0) \rceil) \quad (3.4)$$

Noteworthy is that expression (3.4) shows that the lengths of each edge synchronize with the radius of the ripples depending on frequency f of ripples.

3.6 Angles of Edges

How to decide angle θ is entrusted to the applications.

The angles of edges could be decided from the relation between the parent nodes and each child node. The parameter used for drawing would then be the category attributes of the nodes. Assuming that the content of each node is a news article, categories such as "Society," "Politics," or "Economy" could be obtained from the content of the article. The angle of each edge having child nodes with the same category is set within a certain constant range of the angle.

In the management of the directory and bookmarks, folders (parent nodes) are made in categories division, and the files (child nodes) belonging to the categories in the folder are stored. In this case, each hierarchy of the tree structure is arranged not as a time series but as categories.

In Ripple Presentation, we treat each hierarchy as a time series. The categories mean angles. That is, our layout method can express both the implicit links of leaf nodes (clustering) and the time series.

4 Implementation

4.1 System Architecture

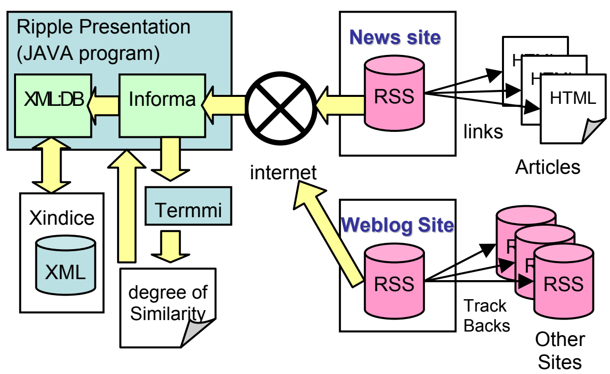


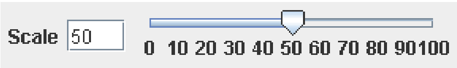
Fig. 3: System architecture.

In our systems, we adopt Java2 SDK 1.4 as the development language and use Xindice1.0 for the native XML database. The Informa library is used for parsing the RSS feed acquisition. If there is no category description in the RSS feed of articles, the degree of similarity among articles is calculated with the external tool "Termmi," so the angle of edge of each article is decided based on this.

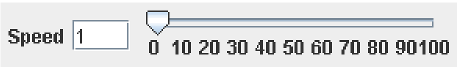
First, the user sets the URL as the location of RSS or trackback URI of the Weblog article to the root node, and the system collects linked information. This information is then converted into XML data, which is registered in the database. Finally, the system draws the graph and ripples based on this XML data in the view. The animation is drawn by synchronizing two threads for ripples and the graph.

4.2 Feature Description

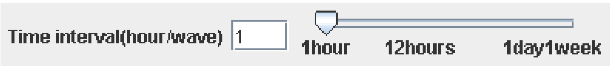
Users can adjust each parameter for an easy-to-view layout while drawing it in the view by animation. As a result, a view is obtained that is well adapted to the characteristics of the data or to the user's viewpoint. Concretely, three parameters are mounted on the upper portion of the view screen and the users are able to control slider (Fig. 4).



(a) Scale Slider



(b) Speed Slider



(c) Time interval Slider

Fig. 4: Sliders for parameter setting.

An example of using these operations is shown in Fig. 5. The viewing time in Fig. 5(b) is 9 days after that in Fig. 5(a). Figure 5(c), which has the same view time as Fig. 5(b), shows a view that has been zoomed out and whose time interval has been changed from 1 hour to a day. Thus, various views can be obtained by changing each parameter according to the purpose.

4.2.1 Scale

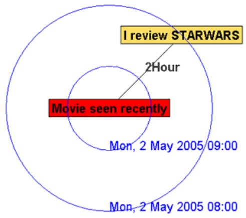
This system has a function that changes the scale parameter of the view. In terms of the model, this parameter reflects the wavelength λ in expression (3.1). Concretely, the view can be controlled using a scale slider to zoom in or out. The default setting of the scale parameter is 50.

4.2.2 Speed

The second function is a controller for the progress speed of the view. This parameter is proportionality constant for progress speed of time t in description of section 3. The default setting is 1.

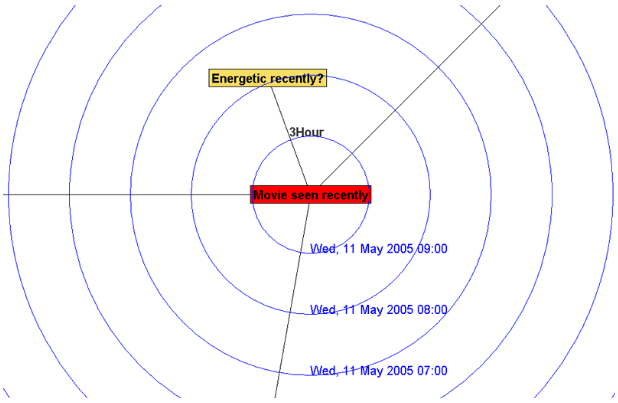
4.2.3 Time interval of Ripples

The third function is a controller for the time-interval of each ripple. This parameter is reflected in the frequency f of ripples in expression (3.1). The initial setting for this parameter is 1 hour (for instance, if a certain ripple is "9 May 2005 10:00," the following ripple is "9 May 2005 11:00," that is after one hour). The time-interval of ripples can be adjusted from 1 to 24 hours (1 day) by 1 hour and from 1 day to 7 days (1 week) by one day.



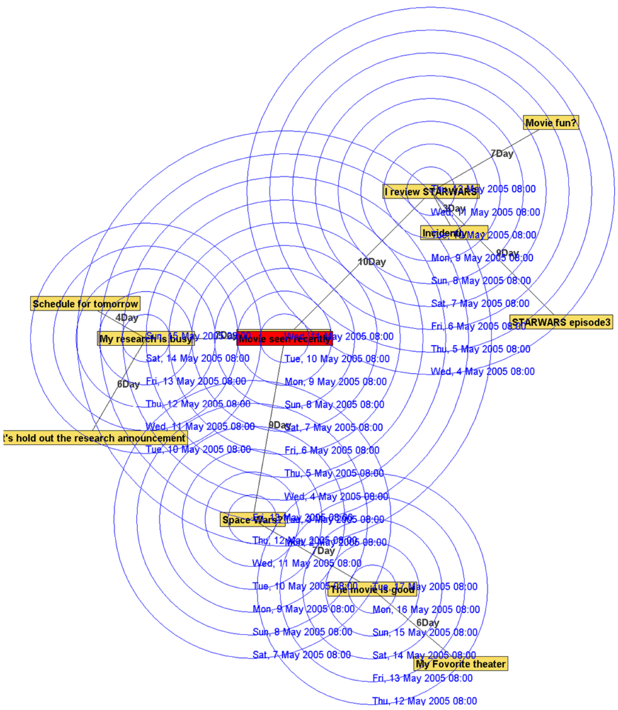
(a) Scale: 50. Time-interval: 1 hour.

(View-time: 2 May 2005 9:00 AM)



(b) Scale: 50. Time-interval: 1hour.

(View-time: 11 May 2005 9:00 AM)



(c) Scale: 25. Time-interval: 1 day.

[View-time: the same as (b)]

Fig. 5: Control flow of parameter setting.

5 Application Example

To demonstrate the utility of Ripple Presentation, we applied it to news articles on a news site, which simply have subtree structure with historical information. In this case, it is possible to decide the graph by comparatively simplifying the angles, because the category is assigned to each article beforehand as an attribute.

We also applied it to trackback articles on the Weblog site, which have rooted tree structure (composed by nested subtree) with historical information. In this case, angle of each edge is decided by the relative similarity between each article after analyzing its contents. Note that ripples are drawn recursively in this case.

5.1 Visualization of Articles on News Site

The information about the latest news with RSS is delivered every day on the News site. Users can watch the latest news or update information easily using the News Reader to collect RSS feeds.

The News Reader often displays news articles using a list with a time series. This presentation method is seen in various scenes, such as Mailer and file management. In general, the users see the titles of articles in a list and searches out a target articles.

In Ripple Presentation, it is possible to refine the search for the latest on-demand news because the articles are categorized by content before users open them in the browser. In addition, we also think that this expression method is useful in analyzing not only the latest article, but also the trends in the news during a certain period and the frequency of the news.

We used RSS (217 articles in total) delivered by the news site "U.S. News and World Report[1]" on September 26, 2005 as data.

The news articles visualized by the Ripple Presentation are shown in Fig. 6. In this case, the root node represents the homepage of the site, and each node is a news article. Each article is categorized by content.

Let us assume that the users want to see articles related to the culture category from the news delivered between 0:00AM, September 11, 2005 and 3:00 AM September 14, 2005. At this time, the users should pay attention to articles appearing in the direction of the angle of the cultural category. In the first step, Fig. 7(a) shows the view where the time interval of the ripples is one week. In the next step, when the time interval of the ripples will be changed to 12 hours, and the root node neighborhood is zoomed, Fig. 7(b) will become the current view. That is, in order to view the latest article from those that appeared during that week, one simply chooses the article nearest the root node and in the culture category.

5.2 Angles of Edges

In Ripple Presentation of the news site, the angle of each edge are divide up evenly in 360 degrees by the number of articles. In addition, all nodes are arranged by category and then, the nodes included every category are arranged by time-series. Therefore, all nodes are sequentially arranged by the time series in the range of the angle of each category.

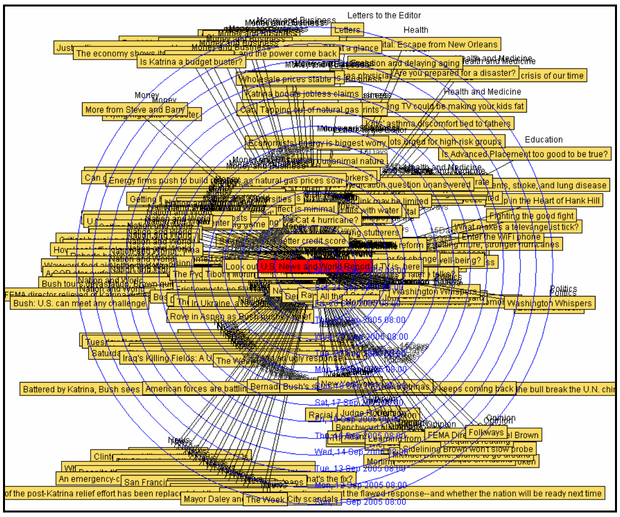
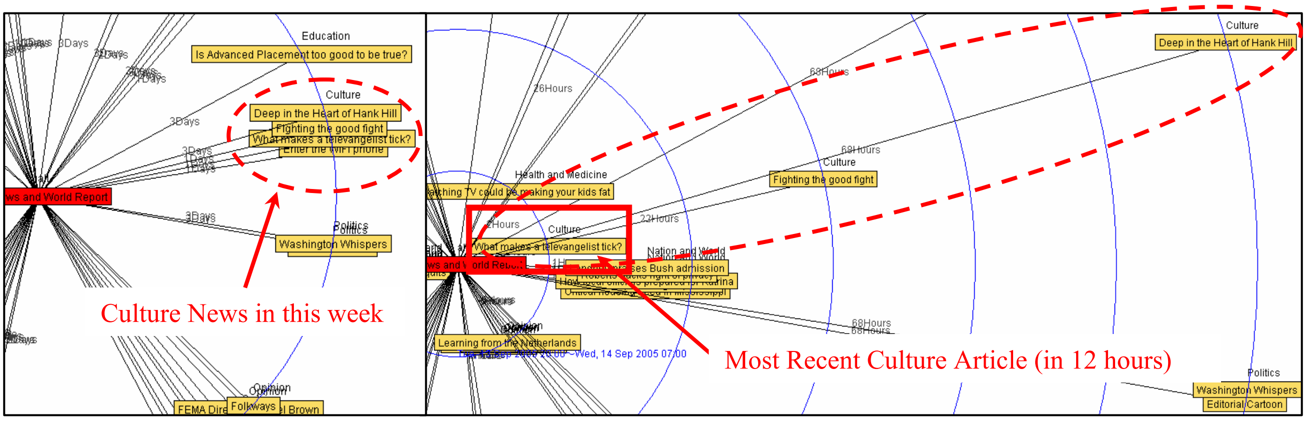


Fig. 6: News list with Ripple Presentation.



(a) Time-interval: 1 week **(b) Time-interval: 12 hours**

Fig. 7: Discovery of latest news articles in the culture category.

[1] http://www.usnews.com/usnews/usnews.rss

5.3 Visualization of Trackback on Weblog

In a nutshell, trackback was designed to provide a method of notification between Weblog sites: it is a method by which person A says to person B, "This is something you may be interested in." To do that, person A sends a *trackback ping* to person B. The URL of the article is described in the RSS feed of the Weblog site of person B by this *trackback ping*. Although the way of understanding the links of the trackback can do nothing but trace the link up to the present time, Ripple Presentation visualizes these links based on RSS.

The structure of the trackback links is different from structure of news articles. The trackback has continuous branching links. That is, the ripples are recursively drawn, because the link structure is a rooted tree.

The links of the trackback to the Weblog article "Google launches Blog Search1" are shown using Ripple Presentation in Fig. 8. The display period was from September 14 5:00 AM to September 18 5:00 AM. The total was 26 articles (until September 30, 2005). Overlapping articles were excluded.

5.4 Angles of Edges

When the trackback of Weblog is visualized, it is necessary to calculate the angles of edges from the relation between each trackback article and the parent article. The procedure is as follows.

1. An important key word is extracted by the *TF/IDF (Term Frequency / Inverse Document Frequency) method* from each trackback article including the parent article.
2. The score of the importance of the article is calculated by the *Vector Space method* from this important key word.
3. The relative difference of the importance score is defined as the degree of similarity between each trackback article and the parent article, and it corresponds to the angle of each edge.

We use the text-mining tool "Termmi" for this purpose (Fig. 3).

If the degree of similarity between the parent article and the trackback articles (child articles) is high, the edges move to the top of the screen. Oppositely, if the degree of similarity is low, they move to the bottom of the screen. Moreover, from the point of view of the parent article, the articles that include unique content are shown on the right side of the screen and those of general content are shown on the left side (Fig. 9).

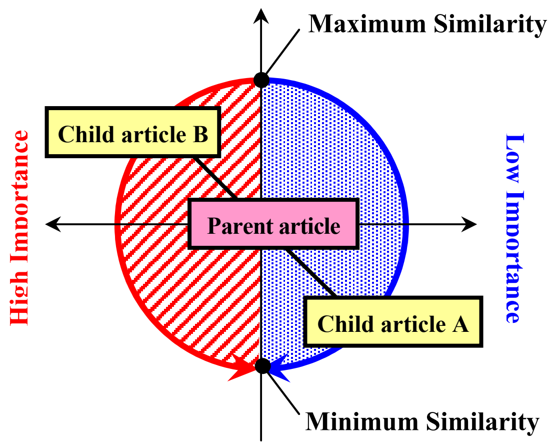


Fig. 9: Angles of edges.

.

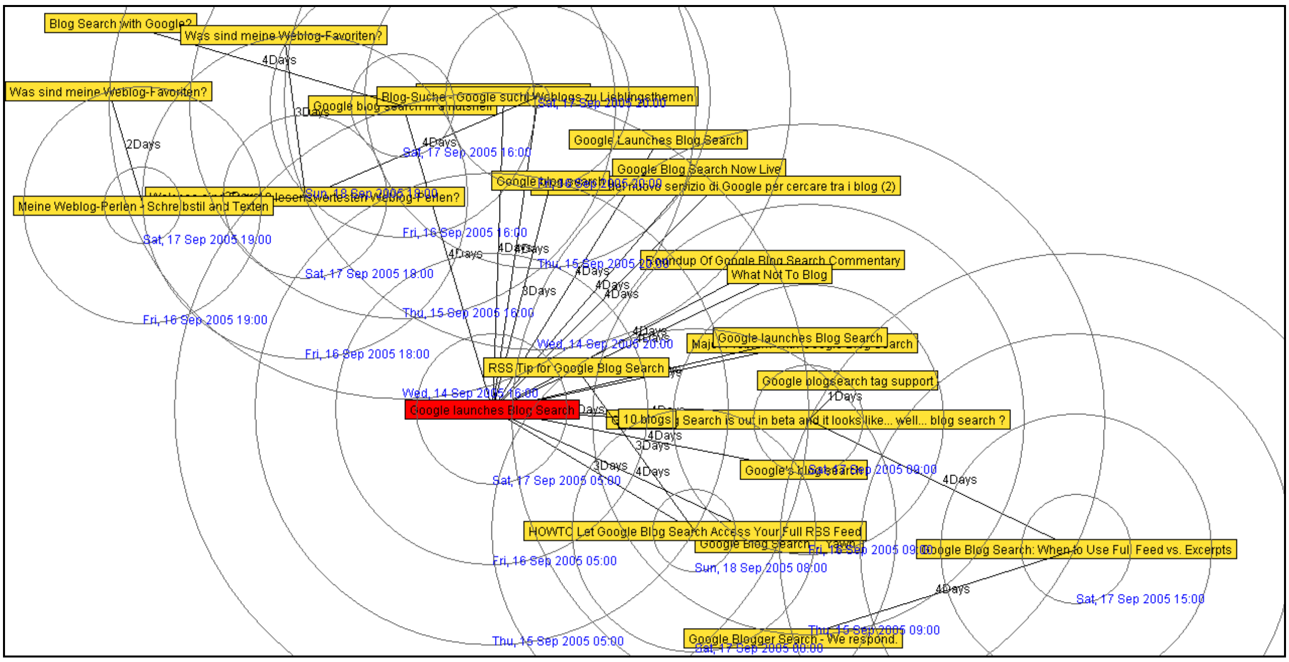


Fig. 8: Trackback links of Weblog articles "Google launches Blog Search[2]" with Ripple Presentation.

---

2 http://www.sixapart.com/pronet/weblog/2005/09/google_launches.html

6 Related Work

A lot of research has been performed for the visualization of the tree structure and many systems have been developed.

A typical tree structure expression technique is the Hyperbolic Tree (Lamping, J., Rao, R. 1993), where the tree diagram is arranged on a hyperbolic space. A node positioned far from the center is displayed smaller, while a node positioned near the center is displayed bigger. This expression method considers the aspect to which the user is paying attention.

Our Ripple Presentation also considers the user's viewpoint. Concretely, the root node is positioned in the center, and the latest node is arranged in the root node neighborhood.

Other research has examined other hierarchies of the tree structure. Tree-Maps (Johnson, B. and Shneiderman, B. 1991) uses space-filling recursive division of a rectangular area as a layout method. Information Cube (Rekimoto, J. 1993) is a visualization technique constructed with a three-dimensional space as a nested construction and features a translucent display of the layered structure. That is, the fundamental purpose of this method is hierarchical clustering for the tree.

On the other hand, our Ripple Presentation is effective when the targeted tree structure data is expressed in a time-series hierarchy. The time-series hierarchy is expressed by concentric circles. Thus, there exists other research to devise the arrangement of the nodes paying attention to the time of the nodes. John V. Carlis focuses on the cycle of time and studied an expression technique for arranging time series information on a spiral as the time-series hierarchy (Carlis, J. V., Konstan, J. A. 1998).

Various algorithms that use concentric circle placement as the drawing convention (like Ripple Presentation does) have been proposed (Wills, G. J. 1999). Eades presented a linear tree-drawing algorithm using a concentric circle layout and proved that the drawing rule of "no crossing" will always be achieved (Eades, P. 1991). Manning and Atallah proposed an algorithm for concentric circle layout that clearly shows "symmetry" (Manning, J. and Atallah, M. J. 1989). We focus on semantic rules, by which users are made aware of the novelty of information by the length of edges and made aware of the categories of information by the angles of edges, rather than by structural rules.

Ka-Ping Yee also expresses each hierarchy of the tree structure in the radial layout. The feature is that the layout is changed by animation dynamically according to user's attention nodes (Yee, K. -P. et al. 2001). Our approach is, dynamically speaking, a circle arrangement based on historical information of nodes rather than rearrangement for clear layout of the tree structure.

prefuse (Heer, J., Card, S. K., and Landay, J. A. 2005) is a toolkit for devising the layout through various effects of animation. The target data is a general graph structure instead of a tree structure. prefuse gives importance to the dynamic change of the layout of the graph based on the user's viewpoint. We also emphasize the user's viewpoint. In Ripple Presentation, the graph is dynamically changed by animation, and the latest information always appears in the neighborhood of the parent nodes.

7 Future Work

There is room for improvement of implementation, such as enhancement of the interface and the data- collection function.

For example, when a category is described for each article like in the RSS of a news site, we want to present to users the range of the category in the view explicitly. By implementing the reading RSS feed of two or more news sites and merging articles, we want users to be able to compare articles about the same news from various viewpoints. It is also important that the drawing of the ripples does not obstruct the user's view. Concretely, it is necessary to devise a system that draws ripples only in the subtree part corresponding to the mouse's position from the user's point of view.

It is necessary to prevent node overlapping, since child nodes closer to a parent node are placed in a more crowded way. An effective solution might be to use the fish-eye lens approach (Lamping, J. and Rao, R. 1993, K. Misue et.al 1995) to zoom the neighborhood of the parent node into focus. Moreover, the crossing of the edges should be reduced to improve readability in Ripple Presentation. It is necessary to model the calculation for angles of edges in designing the layout.

8 Summary

To support efficient information search by presenting tree structure data with historical information visually, we proposed the Ripple Presentation as a visualization technique. This method puts the focus on the expression of the novelty of information, which is difficulty to visualize using a tree-structure expression like a directory.

The advantage of this expression method is that it becomes easy to search for the latest information, because the latest information is arranged in the neighborhood of parent nodes by making a correspondence between the length of the edges and the elapsed time. Moreover, the view in the categorization graph is obtained by making a correspondence between the angle of an edge and a category or similarity. As a result, the users focus attention only to the node of the category of interested.

As an application example, articles on a News site and trackback on a Weblog were expressed using Ripple Presentation. In both cases, the value lies in the novelty of the article. In addition, we can obtain not only the latest information, but also, from an overview of the article group, we can analyze trends in the article group, the appearance cycle, frequency, and so on.

The Ripple Presentation is applicable only for tree structure data with historical information. However, it seems that the range of applications of the Ripple Presentation is wide. In the real world, quite a lot of data with historical information exists. In this paper, we used RSS as target data, which is available as well-structured data resource in spite of a large amount of information. We hope that Ripple Presentation will be used in the various areas of information visualization in the future.

References

Carlis, J. V., Konstan, J. A. (1998): Interactive Visualization of Serial Periodic Data. *Proc. the 11th annual Symposium on User Interface Software and Technology*: 29-38.

Eades, P. (1991): Drawing Free trees. Res. Rep. IIAS-RR-91-17E, Intern. Inst. For Advanced Study of Social Information Science, Fujitsu Lab. Ltd.: 1-29.

Heer, J., Card, S. K., and Landay, J. A. (2005): prefuse: a toolkit for interactive information visualization. *Proc. of Human Factors in Computing Systems 2005*: 421-430.

Johnson, B. and Shneiderman, B. (1991): Treemaps: A Space-Filling Approach to the Visualization of Hierarchical Information Space. *Proc. of the 2nd International IEEE Visualization Conference*: 275-282.

Lamping, J. and Rao, R. (1996): The Hyperbolic Browser: A Focus+context Technique for Visualizing Large Hierarchies. J*ournal of Visual Languages and Computing* **7**(1): 33-55.

Manning, J. and Atallah, M. J. (1989): Fast detection and display of symmetry in trees. Congressus Numerantium.

Misue, K., Eades, P., Lai, W. and Sugiyama, K. (1995): Layout Adjustment and the Mental Map. *Journal of Visual Languages and Computing* **6**(2): 183-210.

Rekimoto, J. (1993): The Information Cube: Using Transparency in 3D Information Visualization. *Proc. of the Third Annual Workshop on Information Technologies & Systems*: 125-132.

Sugiyama, K. (2002): *Graph Drawing and Applications for Software and Knowledge Engineers.* USA, *World Scientific Publishing Co. Pte. Ltd*.

Termmi: Text Mining Shareware Software, http://gensen.dl.itc.u-tokyo.ac.jp/termmi.html, Accessed 27 September 2005.

Wills, G. J. (1999): Niche Works - Interactive Visualization of Very Large Graphs. *Journal of Computational and Graphical Statistics* **8**(2):190-212

Yee, K. -P., Fisher, D., Dhamija, R. and Hearst, M. (2001): Animated Exploration of Dynamic Graphs with Radial Layout. *Proc. of the IEEE Symposium on Information Visualization 2001*:43.