

# Navigating Cyberspace with CyberGeo Maps

Lars Erik Holmquist<sup>a</sup>, Henrik Fagrell<sup>b</sup> and Roberto Busso<sup>c</sup>  
leh@informatics.gu.se<sup>a</sup>, fagrell@informatics.gu.se<sup>b</sup> and  
roberto@informatics.gu.se<sup>c</sup>

The Internet Project<sup>ab</sup> and Viktoria Research Institute<sup>abc</sup>  
Department of Informatics<sup>ab</sup>, Göteborg University  
S-405 30 Göteborg, SWEDEN

## ABSTRACT

*This paper describes CyberGeo Maps, a novel method for visualizing World Wide Web sites. The visualization gives an indication of a web site's size, age, and directory structure. Two examples are shown, one based the site of a large Swedish newspaper, the other on the site of the Association for Computing Machinery. We argue that maps such as these may help researchers and practitioners understand the geography of cyberspace.*

**Keywords:** World Wide Web, information visualization, geography of cyberspace, maps, navigation

**BRT Keywords:** AB, CB, GB

## INTRODUCTION

### The Importance of Maps

In our daily life we often refer to maps of physical spaces, such as cities or buildings. As we spend more and more time online, in “cyberspace,” it seems natural to seek to make maps that chart this virtual space. Many attempts to construct such maps have been made (see e.g. Dodge, 1998). However, it is important to remember that maps are more than just an aid for finding the way from one point in the world to another. Maps are also things of aesthetic and cultural value, which show visions of the world as imagined by the human mind. The map does not necessarily need to be completely accurate, since its only helps its users in gaining a mutual understanding utilized by a common ontology. Maps also can be historical documents that help us understand the world-view of people of other ages. In the future, maps of cyberspace might fill the same function for researchers that ancient geographical maps do today.

### The CyberGeo Project

The point of departure for the CyberGeo project was to measure and monitor the World Wide Web using semi-automatic techniques (Fagrell & Sørensen, 1997). The intention

was to gain an understanding of how the web is developing, which would be of great interest both for researchers and web professionals. To accomplish this a sample of sites were selected and then monitored over a period of time, e.g. six months, using a web robot (Koster, 1998). However, the amount of data generated was enormous, and we realized that we would need information visualization techniques to explore the data effectively. This led us to constructing CyberGeo Maps.

## Related Work

There are several efforts that have explored how to map web hypertext structures by performing analyses of link topologies between web sites (Drew *et al.* 1995; Girardin 1996; Mukherjea & Foley 1995). Carrière & Kazman (1997) have visualized results from searches on the web based on connectivity and content. Bray (1996) applies statistical analyses of the volume and density of the web. To support navigation Dömel (1994) has developed techniques based on graph theory and Lagus *et al.* (1996) self-organizing maps based on neural computing. The most recent survey that we have found is Dodge (1998), who presents a comprehensive list of references.

In the field of information visualization, many methods have been proposed to visualize complex hierarchical structures such as web sites. The *Hyperbolic Tree Browser* applies a view inspired by hyperbolic geometry to present hierarchical web structures (Lamping *et al.* 1995). *Cone Trees* allows visualization of hierarchical structures by presenting them as trees in a 3-dimensional space (Robertson *et al.* 1991). *Tree-maps* is a space-filling tile-based visualization method that has been used to visualize directory structures (Johnson & Shneiderman 1991). *PAD++* is a zooming user interface, which can be used when browsing the web (Bederson & Hollan 1994). The *Zoom Browser* gives users an overview of one or more web pages and might be extended to visualize entire web sites (Holmquist 1997). A combination of the 3D-based methods *Time Tubes* and *Disk Trees* have been used to visualize dynamically changing “web ecologies” (Chi *et al.* 1998).

## CYBERGEO MAPS

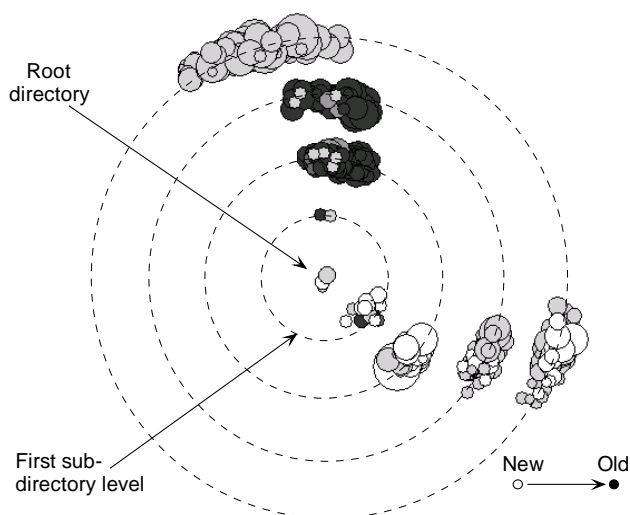
We wanted the web site maps produced by the CyberGeo project to be pleasing to the eye. We also wanted them to show a unique view of some of the “geographical” features of a web site. For these reasons, we let ourselves be inspired by the imagery of space exploration. Drawing inspiration from pictures of planets, galaxies, solar systems and asteroid fields, we imagined a web site as a “sun,” with celestial bodies in the form of web pages orbiting around it.

In the actual visualization, this was accomplished in the following way: Each document on a web site was represented as a dot, with the index page as center. The size of the documents was reflected in the size of the dots. The age (which was defined as the last modification date) was indicated by the shading, so that older documents were shown as darker dots. Dots were then placed at a distance from the center based on the corresponding document’s depth in the file structure of the server, so that dots representing documents one directory deep were placed at distance  $d$  from the center, documents two directories deep were placed at distance  $2d$  and so on. A simple hash function based on the name of the directory and the document distributed the dots in a circle around the center. The hash function was constructed in such a way that documents

in the same sub-directory would occupy approximately the same area, the motivation being that directory structures have meaning since they entail a measure of categorization (Bowker & Star, 1991). Finally, a small element of randomness was added to distribute the dots over a larger area.

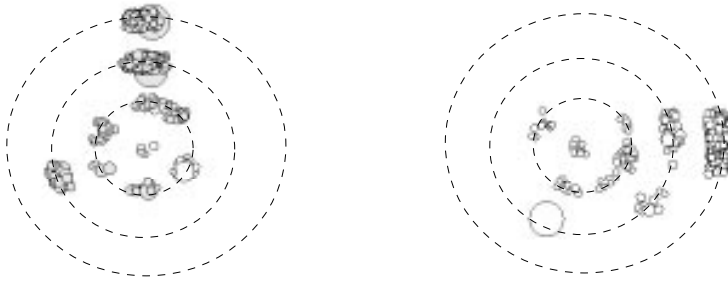
In practice, the visualization was accomplished by first writing a converter program in Java, which took as input the log-files, generated by the web robot, and generated a database containing all necessary information. The database was then input into Spotfire (<http://www.spotfire.com/>); an interactive visualization tool based on the concepts of dynamic queries (Ahlberg & Shneiderman 1994). By tight coupling of query devices (e.g. range-sliders and checkboxes) to a scatter-plot display, the visualization tool allows for interactive exploration of a data set, including zooming and searching.

This visualization method provided us with a view that reflected the structure, size and age of a web site, and that came reasonably close to our original vision of “astronomical” maps. Sites that were new or frequently updated might appear as solar systems with bright suns, whereas old sites would appear as “black holes.” Extending the metaphor further, it would be possible to construct “star-maps,” consisting of many web sites, where the age and size of each site would be apparent at a glance.



**Figure 1: Explaining CyberGeo maps.**

Figure 1 shows an example of a map generated in this way. Each dot in the figure represents a document on the server. The relative size of the documents can be seen from the size of the dots, and the age is reflected by a gray-scale coloring. The collection of dots in the center represents the index page of the site (i.e. the page which is shown if only the server name is indicated as address, e.g. <http://www.gp.se/>) and a few other pages placed in the root directory of the server. The clusters of dots placed in concentric rings around the center represent documents placed in various directories on the server. Documents in the same directory are placed close together, forming the clusters, which can be seen in the figure. Since this placement carries over into the next directory level, relations between directories and subdirectories can easily be seen in the figure.



**Figure 2: An example of radical change.**

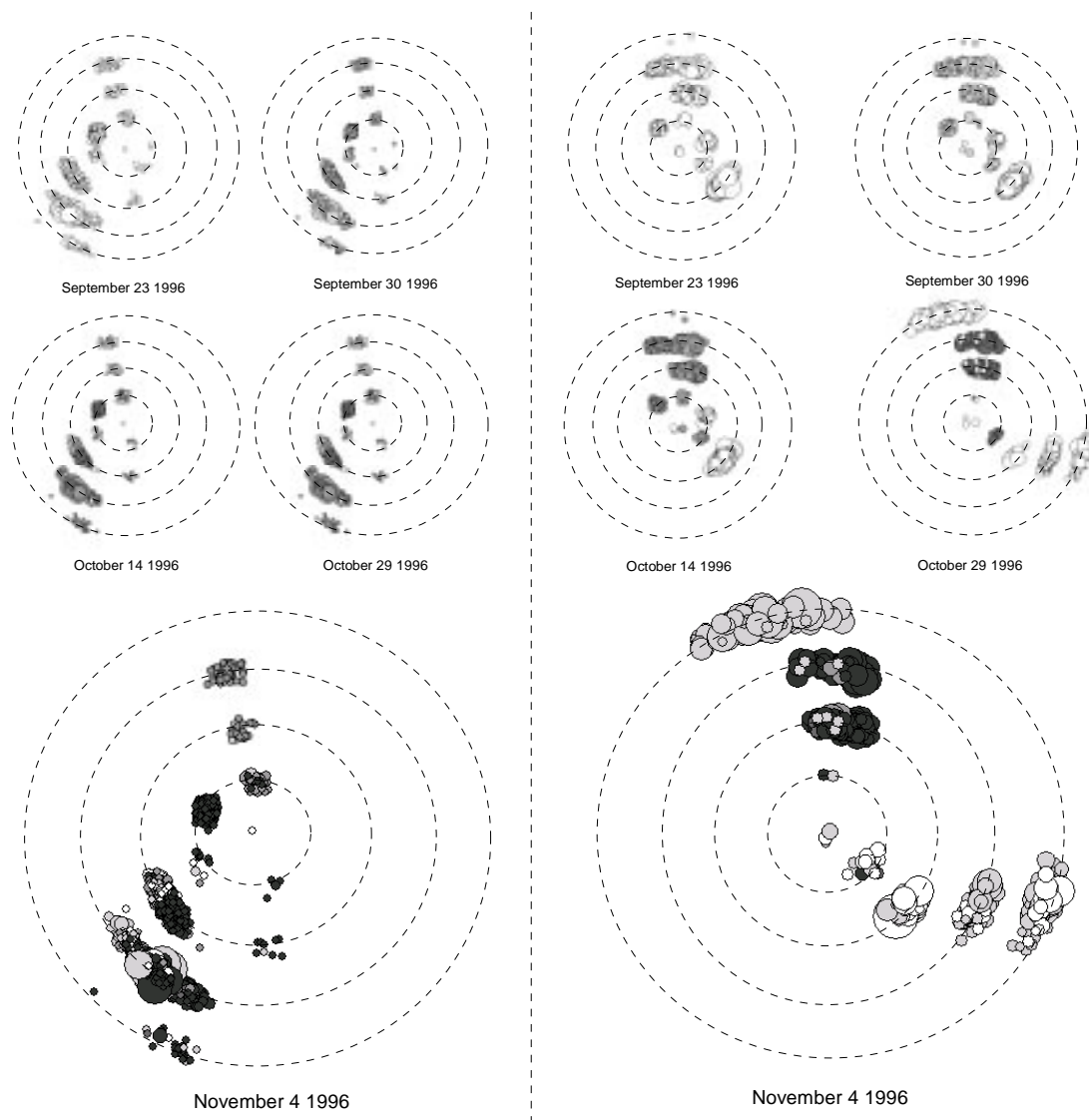
Figure 2 gives a practical example of how the visualization can be used to monitor a web-site over time. The figure shows how the organization of a web-site has changed drastically between two occasions. Some changes, which can be seen in the figure, are:

- The number of pages in the root directory has approximately doubled in the second image
- A single, large documents has been added or moved to a 2<sup>nd</sup> level subdirectory
- Although the number of 1<sup>st</sup> level subdirectories is the same, two of them have been renamed
- The contents of the 3<sup>rd</sup> level subdirectory has changed drastically, with one large document removed and a large number of smaller documents added
- A large documents in a 2<sup>nd</sup> level subdirectory has been removed or moved to another directory

More details can be found by exploring the visualization in the visualization tool where zooming is possible. The important thing is to see how the visualization makes it possible to identify many small and large changes in a web-site at a glance - changes which may require much more work to find by studying or processing web robot log files.

## RESULTS

In the following, we will discuss how CyberGeo Maps have been used to visualize data collected in the CyberGeo project.



**Figure 3: Web sites of two Swedish newspapers: Sydsvenska Dagbladet (www.sds.se) and Göteborgs Posten (www.gp.se ).**

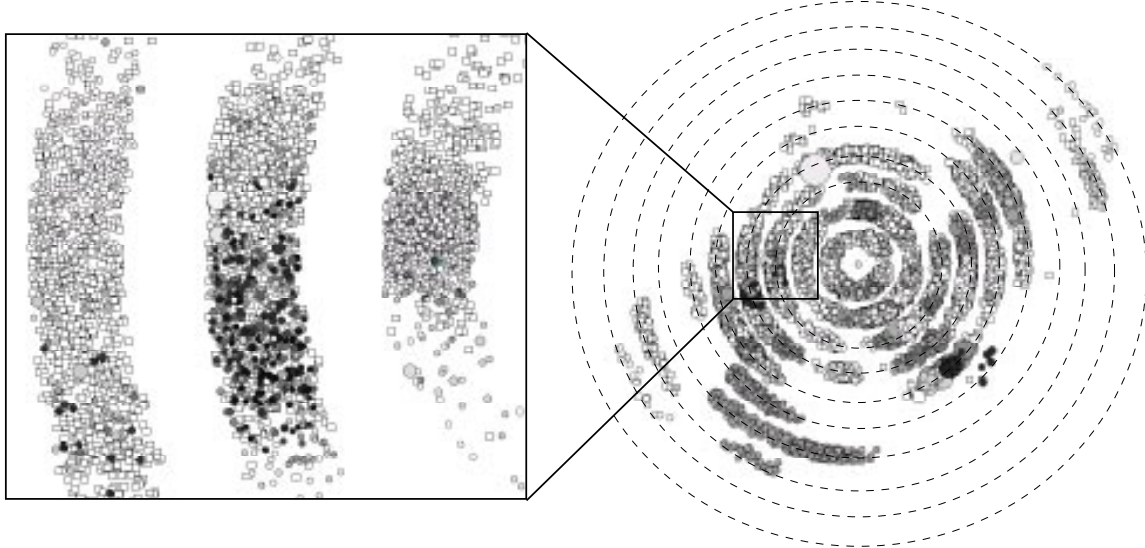
### Example 1: Two Swedish newspapers

Figure 3 shows a visualization of data collected in the CyberGeo project. It shows the development of the web sites of two of Sweden's largest newspapers, Sydsvenska Dagbladet (SDS) and Göteborgs Posten (GP), during the course of a few weeks. The web sites were of approximately equal size, containing about 1,000 documents each. In this example, the shading was based on a relative time-scale, so that all documents were classified as "new" during the first run, and documents that remained unchanged grew progressively darker.

The SDS site shows few dramatic changes, and the only page that seems to be consistently updated is the index page (the central dot is white in all diagrams, indicating that it has been changed since the last occasion). A number of documents are added or updated for each measure point, but a large number of documents also remain completely unchanged during the entire period. No new directories have been added during the period of measurement.

The GP site shows more changes, and even a brief inspection shows that it is

more frequently updated than the SDS site (it is generally “lighter” in color, containing a higher number of recently updated pages). Note that the 2<sup>nd</sup> and 3<sup>rd</sup> subdirectories situated at approximately 12 o’clock in the diagram seem to go almost unchanged during the entire period, with only a few documents being added or changed. An interesting event occurs between the measure points of October 14 and October 29, when a large cluster of documents is added to the 4<sup>th</sup> subdirectory between 11 and 12 o’clock. This cluster of documents is completely unchanged in the next measure point, indicating that a large amount of mainly static material was added to the site at this time.



**Figure 4: The web site of the Association for Computing Machinery (www.acm.org).**

## **Example 2: The Association for Computing Machinery**

In Figure 2, 17,000 pages collected from the web site of the Association for Computing Machinery on December 11, 1997, were visualized. This is a very large site, and the complexity of such a large material is clearly indicated by an enlarged portion of the map. The shading in this case was based on an absolute time-scale, defined as a document’s age in days from January 1, 1993. (Although older documents did exist on the site, they were very few.) Documents for which no age information could be obtained were indicated as unshaded squares.

Examining the diagram shows that the site has a fairly deep directory structure (ten levels of subdirectories) but that most material seems to be concentrated in the lower levels. It is easy to find “outliers” among the documents with simple visual inspection. Extremely large documents are indicated as large dots (see subdirectory level 4, at approximately 11 o’clock, where a very large postscript file is situated) and areas which are very old (see e.g. subdirectory level 5 and 6 at approximately 4 o’clock, which is an archive area belonging to the SIGCOMM group).

## **FUTURE DIRECTIONS**

Many improvements can be made to the maps presented here, both aesthetically and functionally. The use of hash functions based on directory names is not necessarily the best method to base the distribution on, and it may be enhanced or replaced with other

methods, e.g. semantical methods. By combining the structure of a web site with log-files of actual use, more information might be displayed, such as the popularity of certain parts of a web site.

Recently, we have tried out the concept of CyberGeo maps to support navigation on the web, with the focus on situations where the screen size is small. We find this important since there is a trend in providing web access on mobile devices (Glave 1997) on which the screen size probably will remain limited for some time. We have looked into complementing techniques, e.g., an autonomous interface agent similar to Letizia (Lieberman, 1997) and Flip zooming techniques (Holmquist & Björk, 1998). Figure 5 shows our first mock-up that we are planning to evaluate.



**Figure 5: A mock-up example on a CyberGeo map with an interface agent in a Flip Zoom container.**

## SUMMARY

The function of maps in cyberspace is still unclear. We have shown one vision of how to map a web site; there are many others. However, our maps seem to work well in giving an overview of the complexity and organization of a web site, without cluttering the view with unnecessary detail. They may be of use for constructors of web sites, who want to make themselves a picture of their site or to anyone who wants navigation support on a mobile device.

In the future, maps such as the CyberGeo Maps may be of importance to historians wanting to understand the development of the World Wide Web.

## REFERENCES

Ahlberg, C. and Shneiderman, B. Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. In Proceedings of CHI'94, ACM Press, New York,

- 1994, 313-317.
- Bederson, B., Hollan, J. PAD++: A Zooming Graphical Interface for Exploring Alternate Interface Physics. In Proceedings of ACM UIST '94, ACM Press, New York, 1994, 17-26.
- Berners-Lee, T., Cailiau, R., Luotonen, A., Nielsen, H.F. and Secret, A. The World-Wide Web. Communications of the ACM 37 (8), 1994, 76-82.
- Bowker G. and Star, S.L. Situations vs. Standards in Long-Term, Wide-Scale Decision-Making: The Case of the International Classification of Diseases. In Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System Sciences, Kauai, Hawaii, January 7-11, 1991, ed. Jay F. Nunamaker, Jr. and Ralph H. Sprague, Jr. IEEE Computer Society Press, vol. IV.
- Bray, T. Measuring the Web. In Proceedings of the 5<sup>th</sup> International World Wide Web Conference. Paris, France, 1996.
- Carrière, J. and Kazman, R. WebQuery: Searching and Visualizing the Web through Connectivity. In Proceedings of the 6<sup>th</sup> International World Wide Web Conference, Santa Clara, California, USA, 1997.
- Chi, E.H., Pitlow, J., Mackinlay, J., Pirolli, P., Gossweiler, R. and Card, S.K. Visualizing the Evlotuion of Web Ecologies. In Proceedings of CHI '98, ACM Press, New York, 1998. 400-407.
- Dodge, M. The Geography of Cyberspace. University College London, UK.  
<http://www.cybergeography.org/>, 1998.
- Dömel, P. WebMap - A Graphical Hypertext Navigation Tool. In Proceedings of the Second International WWW Conference, Chicago, USA, 1994.
- Drew, N.S. and Hendley, R.J. Visualising Complex Interacting Systems. In Proceedings of CHI'95. ACM Press, New York, 1995, 204-205.
- Fagrell, H. and Sørensen, C. It's Life Jim, But Not as We Know It! In Proceedings of WebNet '97, Toronto, Canada, 1997.
- Girardin, L. Mapping the virtual geography of the World-Wide Web. In proceedings of the 5<sup>th</sup> International World Wide Web Conference. Paris, France, 1996.
- Glave, J. Handheld Internet Will Be Huge - Really! Wired News.  
<http://www.wired.com/news/news/technology/story/8234.html>, 1997
- Holmquist, L.E. Visualizing Web Pages with Flip Zooming and the Zoom Browser. In CHI '97 Conference Companion, ACM Press, New York, 1997.
- Holmquist, L.E. and Björk, S. Formative Evaluation of Flip Zooming: Towards Effective Integration of Detail and Context on Small Displays. Technical Report VRR-98-1, Viktoria Research Institute, Göteborg, Sweden, 1998.
- Johnson, B and Shnedierman, B. Tree-maps: A space-filling approach to the visualization of hierarchical information structures. In Proceedings of IEEE Visualization '91, IEEE, Piscataway, NJ, 1991.
- Lagus K., Kaski, S., Honkela, T. and Kohonen, T. Browsing digital libraries with the aid of self-organizing maps. In Proceedings of the 5<sup>th</sup> International World Wide Web Conference, Paris, France, 1996.
- Lamping, J., Rao, R. and Pirolli, P. A Focus+context technique based on hyperbolic geometry for viewing large hierarchies. In Proceedings of CHI '95, ACM Press, New York, 1995, 401-408.
- Lieberman, H. Autonomous Interface Agents. In Proceedings of CHI-97, ACM Press, New York, 1997, 67-74.
- Mukherjea, S. and Foley, J.D. Visualizing the World-Wide Web with the Navigational View Builder. Computer Networks and ISDN System, Special Issue on the 3<sup>rd</sup> International Conference on the World-Wide Web, April 1995, Darmstadt, Germany.
- Koster, M. The Web Robots Pages. <http://info.webcrawler.com/mak/projects/robots/robots.html>, 1998.
- Robertson, G.G, Mackinlay, J.D. and Card, S.K. Cone Trees: animated 3D visualizations of hierarchical information. In Porceedings of CHI '91, ACM Press, New York, 1991, 189-194.