Toward Visual Web Mining

Amir H. Youssefi[1], David J. Duke[2], Mohammed J. Zaki[1], and Ephraim P. Glinert[1]

[1] Department of Computer Science,
Rensselaer Polytechnic Institute,
110 Eigth Street, Troy, NY 12180
{Youssefi,Zaki,Glinert}@cs.rpi.edu
http://www.cs.rpi.edu/~youssefi/research/VWM
[2] Department of Computer Science,
University of Bath,
Bath, BA2 7AY, U.K.
D.Duke@bath.ac.uk

Abstract. Researchers analyzing web sites face two challenges of the increasingly large amount of data published online as well as web sites with more complex structures. We apply Data Mining and Information Visualization techniques to the web domain in order to benefit from the power of both computing and the human visual perception; we term this Visual Web Mining. In response to the two aforementioned challenges, a generic framework is proposed. Within this framework, we apply Data Mining techniques to large datasets and use Information Visualization methods on the results. The goal is to compare the outcomes of mining Web Usage Logs and extracted Web Structure by visually superimposing the results. We propose several new information visualization diagrams and analyze their utility, elaborate the architecture of a sample implementation, and present work in progress, along with early results and discussion of ideas for future work.

1 Introduction

Analysis of web site regularities and patterns in user navigation is getting more attention from businessmen and research community as web browsing becomes an everyday task for more people around the world. Lots of efforts has been made for exploring these regularities and some through visualization [6, 7, 10, 14, 19, 34].

To enable visual superimposition, we have designed new visual representations of site structure and user access patterns. The second part of this paper discusses the ideas behind these and the implementations.

Aspects of our representations have been motivated by metaphors developed initially within scientific visualization, in particular the representation of vector-fields in fluid dynamics. The results are interactive 3D visualizations implemented in the Visualization Toolkit [29], providing exploratory tools for supporting Visual Web Mining (VWM).

The architecture of our implementation, work in progress and early results are presented and ideas for future work are discussed. We leave further analysis and usability tests for future work.

Visualization is defined as the use of computer-supported, interactive, visual representations of data to amplify cognition [2]. Some surprising examples in which a good visualization can unlock previously hidden structure and yet not obscure the detail are given in [9]. *Information Visualization* is the use of computer-supported, interactive, visual representations of abstract data to amply cognition [2]. The field of *Information Visualization* is about creating tools that exploit the human visual system to help people explore or explain data [25] or simply visualization applied to abstract data [2]. Computer support brings the opportunity to explore radically different representations, including the use of 3-dimensional models.

The capability to interact with a visual representation is significant in allowing users to explore large-scale datasets, where it is infeasible to provide both an overview of the space plus information about points of focal interest (the so called "focus plus context problem").

Different approaches have been taken by researchers to visualize information some notable ones are [8, 11–13, 15–17]. Visual Data Mining [5] is of particular interest.

A key challenge in Information Visualization is finding a spatial mapping for an abstract data set that is cognitively useful for a specific task. To address this, information visualization draws on ideas from several intellectual traditions, including computer graphics, human-computer interaction, cognitive psychology, semiotics, graphic design, and cartography.

Although there is still disagreement on the distinction between Scientific Visualization and Information Visualization, for the purposes of this paper the distinguishing feature of information visualization is taken to be the need to find a spatial mapping of data that is not inherently spatial, whereas scientific visualization uses a spatial layout that's implicit in the data.

Web Mining as application of data mining techniques to the World Wide Web, can be devided into Web Content Mining, Web Usage Mining and Web Structure Mining. In this paper, our focus is on last two parts because of implementations available to us. We define the notion of a *user session* as a temporally compact sequence of web accesses by a user. The goal of our web mining in part is to work on these sessions for better visualization toward useful information.

A common theme underlying the use of visualization in website analysis is the graph metaphor, that is, the organization of a web site and/or patterns of access are treated as a node-link graph. There is a considerable literature on algorithms for drawing graphs however, making *aesthetically pleasing* drawings of graphs (e.g. with a minimal number of edge crossings) is computationally expensive, and for large-scale graphs new techniques have been developed, see [4] for case studies, and [26] for a survey of recent approaches. One issue is that drawing general graphs is harder than drawing trees, for which a number of efficient approaches are known, for example cone trees [8]. Where a graph is not itself a tree, tree layout can be applied to a spanning tree of the graph, with non-tree edges either added to the resulting structure [1], or not included if the spanning structure is sufficient for a given task. For example, the structure of a web site is in general a graph, with pages corresponding to nodes, and links to edges. The high-level organization of a site is often hierarchically i.e. as a tree, with one topic branching into a number of sub-topics, as in linear printed media. However, a strength of hypermedia is the ability to cross-link topics, and also provide navigational paths back to higher-level areas, and these kind of links then form non-tree edges of the corresponding graph.

An interesting variation on the node-link representation of trees is the tree-map representation [28]. In the context of visualizing data mining results, the ability of this representation to convey large-scale datasets is an advantage. The limitation of this approach is that, while it is good at conveying properties of nodes, it does not realiy convey paths or attributes of sub-structures. Nor is it clear whether non-tree structure can be attached to the representation.

2 Overview of Visual Web Mining Framework

We propose *Visual Web Mining* (VWM) as application of Information Visualization techniques on results of Web Minining in order to further amplify perception of extracted patterns, rules and regularities or visually explore new ones on web domain.

We describe a framework i.e. *Visual Web Mining Framework* and use it in a sample implementation and there we construct useful diagrams using Information Visualization techniques on the results of Data Mining Algorithms such as Sequence Mining[36], Tree Mining [35] with input of large web access log files as well as huge Web Graphs. Web graphs are semi-static snapshots of web site structure, in contrast web access logs capture the dynamic behavior of surfers visiting a web site. Visual superimpostion of dynamic and static views enables web analyzers or even web masters to compare these two aspects and gain insight on what actually happens on a website.

The organization of large websites reflects a diverse range of design criteria and rationale; e.g., personal preferences of web designers, and the breadth and structure of the topics presented on the site. In the second part of this paper we use visualization to obtain an understanding of the structure of a particular website, as well as web surfers' behavior when visiting the site.

Due to the scale of dataset on which we work, and the structural complexity of the sites from which the data is obtained, we have decided to use 3D visual representations, with user interaction to support navigation and exploration. We have implemented this using an open source toolkit called the Visualization ToolKit (VTK) [29].

The decision to implement website visualization using a toolkit developed, at least initially, for engineering and medical applications, deserves further comment. A number of public-domain tools for the visualization of graphs do exist, and some can cope with large graphs, with up to one million nodes. However, these tools typically provide a fixed repesentation of a graph (a node-edge diagram), and have limited modularity. In the case of visual web mining, we wish to experiment with novel combinations of metaphors, including spatial positioning, tubing of edges, node glyphs, and color mapping. These techniques are provided in a modular way within general visualization tools, such as VTK, and we have therefore chosen to extend VTK with specific support for working with node-link graphs. Further rationale and design issues underlying such an approach can be found in [3].

Although the work is at an early stage, and improvements in both mining techniques and visual representation are ongoing, early results already show different notable phenomena e.g. Cone, Funnel and Debris which corresponds to user behaviors as he/she browses the website (see figure 2). Different mappings are made from data attributes to visual attributes on each diagram. Visualization techniques are used to portray mined structures via visual features such as splines, trees, and stream tubes, with visual attributes such as color, and size used to convey interesting data features, for example users leaving the web site.

Implementation of this Visual Web Mining Framework has been applied to the website and access logs of the Computer Science Department of Rensselaer Polytechnic Institute (http://www.cs.rpi.edu).

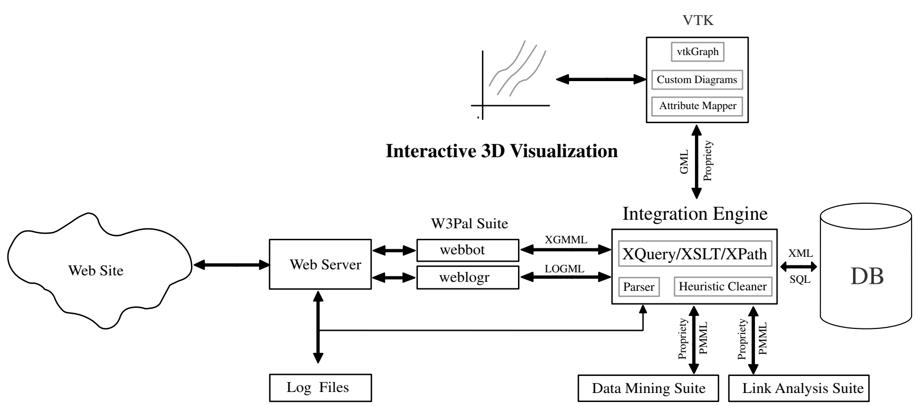

Fig. 1. Sample implementation architecture of VWM

3 Architecture

Figure 1 shows the architecture of our implementation. We target one or a group of websites for analysis. Input of the system consists of web pages and web server log files. Access to web log is either through the local file system, or by downloading it from a remote web server (possibly as a password protected resource).

A web robot (webbot) is used to retrieve the pages of the website and then construct a web graph. Our implementation uses the robot of the WWWPal Suite [32] to generate this graph, which is subsequently converted to the simple GML format used by the visualization tools.

In parallel, Web Server Log files are downloaded and processed through a sessionizer and a LOGML [30] file is generated. Web logs could be in either Common Log Format or Extended Common Log Format. Both input and output of different parts are in standard (XML) languages so output of other similar programs could be fed to the system with small changes also results can be used by other systems.

The Integration Engine is a suite of programs for data preparation i.e. cleaning, transforming and integrating data. It uses XQuery, XSLT and XPath as well as Regular Expression Parsing on XML languages also propriety text data to do the transformation. Support for PMML is readily available as well. A connection module is in charge of Bulk Loading of XML data into database and executing SQL commands against the database. Schema matching is done using external tools as well as code snippets which map different schema and import/export different XML/text files into/from relational tables in our database server. Web access log files are also imported into database for exploring and comparison of Data Mining Algorithms as well as verification of data integration.

Data Mining Suite and also *Link Analysis Suite* need special data format as input and give output in propriety formats hence the Integration Engine is required to covert data in these formats.

The visualization stage of this pipeline, which maps the extracted data and attributes into visual images, is realized through VTK extended with support for graphs. VTK[29] (and this extension) is a set of C++ class libraries, available on a range of architectures including Windows and Linux. The class library is accessible either through linkage with a C++ program, or via wrappings supported for scripting languages (Tcl, Python or Java). In the case of this work, the visualization engine is delivered in the form of a Tcl script. Results are interactive 3D/2D visualizations which could be used by analysts to compare actual web surfing patterns to expected patterns, or more generally, the intended purpose and role of the website.

Web Mining has no single recipe for satisfying requirement of analysts or business managers. Businessmen ask high level, diverse questions[20], for example:

- Is our site sticky? Which regions in it are not?
- How adept is our conversion of browsers to buyers?
- What site navigation do we wish to encourage?
- What attribute describes our best customers?
- What makes customers loyal?
- How can profiling help use cross-sell and up-sell?

These questions are semantically distant from the data available for analysis, and suitable responses are beyond the scope of either simple data mining or visualization. What VWM can provide, however, is insight on more specific, focused questions, for example:

- What is the typical behavior of a user entering our website?
- What is the typical behavior of a user entering our website in page A from 'Discounted Book Sales' link on a referrer web page B of another web site?
- What is the typical behavior of a logged in registered user from Europe entering page C from link named "Add Gift Certificate" on page A?
- What is the typical behavior of a user who come in our website 1 day to 3 weeks before Christmas and buys something, versus one who didn't by anything?

In order to partially support this kind of analysis we take the following approach in our example implementation of the VWM framework:

- Make personalized results for targeted web surfers (users, customers) as opposed to blindly aggregating on all users.
- Use Data Mining Algorithms for extracting new insight and measures.
- Employ a database server and relational query languages as means to submit specific queries against data e.g. projection and aggregations, joins etc.
- Utilize visualization to obtain an overall picture correlating static site structure with (dynamic) access patterns. Support 'at a glance' overviews using visual aggregation and filters. Amplifying special features via interactive control.

A key issue for the data mining part of this process is how one translates the notion of the *typical behavior* of a user into actual queries on data sets to yield sufficient insight for an analyzing team to identify whether (or how) the (business) goals of the website have been satisfied relative to those targeted customers.

These points serve to emphasize how visualization complements data mining. The latter is about utilizing capabilities of the machine to find and/or compute patterns within the dataset, based on notions of pattern derived from domain knowledge or statistics. Visualization, in contrast, is about using human capabilities to detect patterns or discern trends within visual representations.

In our approach we first extract user sessions from web logs, this yields results related to a specific person as much as possible. User sessions are then converted into a special cSPADE [36] format:

```
<customer_id> <time_id> <item1> <item2>
```

Here `item1` is the smaller numerical value of SourcePageID and TargetPageID, while `item2` is the larger numerical value of TargetPageID and SourcePageID; this ordering is needed as input to cSPADE needs to be sorted. This input means that a customer (web surfer) bought `item1` and `item2` (and thus visited the source and target pages) together at her shopping (click).

For each click of user we need to have both source page and target page, not just source or target page, in order to deal with different cases in user navigation such as Hitting Browser's Back Button or finding Entry Page or Exit Page. Furthermore, Source pages (or target pages) out of our website could either be considered as a single page or for more details analysis multiple PageIDs could be assigned by a classification or clustering algorithm say on referrer URL (or Exit Page) to gain more insight.

We convert text input data into binary format, vertical format, make indices and run cSPADE algorithm using lower limit of 1 and upper limit of 1 in order to get *continues* frequent sequences with a given support. Results are imported into databases, shorter frequent sequences are removed. Later different queries are executed against this data according to some criterion e.g. support of the patterns found, length of found patterns etc. This gives the flexibility to translate *typical behavior* into different appropriate queries according to subjective requirement set forth by analysis team without going all the way down to web log analysis again.

Result is a graph rather than a simple sequence since user can go back from one page to previous ones or follow anchors in the same web page or simply follow a closed path as sequence of pages.

We attempt to present abstractions of large amounts of data in tuned diagrams in order to maximize opportunities for pattern detection via human perceptual capabilities. We have designed several visualization diagrams, some of which have been implemented. The contribution of some of those diagrams is discussed and illustrated shortly; video clips of the system in action are available from the webpage of the VWM project [31].

4 From Fluid Dynamics to Web Dynamics

Advances in scientific visualization techniques, particularly in the domain of fluid dynamics [21–23], have inspired us to apply similar techniques in information visualization on the web domain. We use an analogy between the 'flow' of user click streams through a website, and the flow of fluids in a physical environment, in arive at new representations As with most problems of information visualization, our representation of web access involves locating 'abstract' concepts (e.g. web pages) within a geometric space. With this as a base, the following ideas motivate our design of the visualization diagrams discussed in the next section:

- Iconic visualization on extracted features from Data Mining.
- Particle tracking for tracing the path of user navigation.
- Feature tracking/detection, for instance turbulence vortex detection techniques, for identifying a user cycling around a set of web pages. Note that a similar phenomena happens for a user navigating pages and this could be interpreted both as confusion and misled tracking or quite reverse a case where a user is focusing on a cluster which he/she has found interesting.

– Splatting for simplying visualization implementation in dense parts (both 'hot' and 'cold' parts of a graph could be splatted in order to focus users' attention on edges connecting these two regions.
– Fluid event detection on navigation clickstreams to capture special (interesting) events.

In [24] visualization techniques were grouped in three categories:

1. *Global Techniques* give a qualitative, global visualization of the data at a low level of abstraction.
2. *Geometric techniques* extract geometric objects (curves, surfaces, solids) from the data. They can be considered as intermediate-level representations, both with regard to locality and level of abstraction.
3. *Feature-based techniques* extract high-level, abstract entities from the data. The emphasis is on quantification for more precise evaluation and comparison.

The second and third of these are used for our visualization methods.

4.1 Structures

Graphs Much work has been done on two and tree dimensional embedding of graphs, in the form of both algorithms and tools, see for example [1, 18, 32, 33]. We have developed a library under VTK for graph visualization[3] (see `http://www.cs.bath.ac.uk/~djd/graphs.html`). Two approaches to visualizing the output of data mining have so far been implemented using this library. In the first approach, the link analysis is treated as a directed graph. This graph is drawn by obtaining a spanning tree for the underlying graph, laying out this tree (using a variation of the cone tree layout first proposed by Robertson et al [8]) and then re-introducing the non-tree edges. Further attributes obtained by data mining can then be superimposed on the underlying graph structure, for example as variable-width tubes showing particularly strong access paths.The second approach to visualization is to take a spanning tree and use this itself as the main visual element, supported by color mapping edges and glyphing nodes. In either case visual attributes of nodes and edges have different mappings/meanings for each of the diagram designed, the details of which are discussed in the next section.

Stream Tubes Variable-width tubes showing access paths with different traffic are introduced on top of the web graph structure. Here, depending on the type of visualization diagram, particular weights e.g. support of a single click-stream, total sum of support on all click-streams, support extracted from Tree Mining Algorithm [35] can be mapped onto the width (radius) of the stream tube. Color mapping could be used on the number of users leaving the website (or a cluster of these in a zoomed view), a property of the graph structure (such as the Strahler value), or simply the number of hits of some branch.

4.2 Design and Implementation of Diagrams

Building on the intuition and heuristics set out above, we have designed new visualization diagrams; these are summarized below, after which we will consider issues of utility.

- Figure 3 is a visualization of the web graph for the Computer Science department of Rensselaer Polytechnic Institute (http://wwww.cs.rpi.edu).
- Figure 4 is a 3D visualization of web usage aggregation for this site.
- Figure 2 is representation of the same site in which color mapping is used to highlight the mass of surfers scattering in any cluster making a Cone shape and later coming back to main pages of clusters and to the first page of the website making a Funnel. A user session drilling down into a cluster in a form of a semi-directed flying debris is also observed.
- Figure 10 is superimposition of web usage mining results (dynamic user access patterns) on top of the web graph (the static link structure of the web site).

A web master can find out where the load of his website and bandwidth goes by a quick look at this kind of diagrams e.g. figure 4 and click on glyphed nodes which represent web pages to see which page each node represents (see figure 9). Cube glyphs with proportional size are put on top of nodes to make them easily clickable. Glyph size adds another visual dimension that can be used for encoding attribute data. The efficiency of the underlying VTK graphics interface (in turn built on an interface to OpenGL or Mesa) is such that users can easily zoom, pan and rotate these specific diagrams, even on modest hardware (e.g. a 433Mhz Pentium II). Thus users can explore different parts of the sample website i.e. http:/www.cs.rpi.edu as extracted from the weblogs. Strategies for real-time interaction with much larger datasets are already being explored.

Web mining results in patterns of most frequent access which are visualized in 5 as white edges superimposed on the remainder of the figure. Figure 6 extends this approach by using thickness of stream tubes to add a further visual dimesion, in this case encoding how frequent those access patterns are.

In figure 10 we have visualized the static structure of website taken from our webbot (gray-coloured edges) to make a basement on which actual dynamic behavior of users is superimposed with (colored edges). There are about half a million visualized nodes. A webmaster or a web analyzer can easily see which parts of the website are 'cold' parts with low hit and which parts are 'hot' parts with high hit. This also paves the way for making exploratory changes in web site and analyze the changes in user access. For instance a webmaster can change link structure e.g. by adding a link of a cold cluster in first page of website. Another example is change of content e.g. by highlighting existing anchor text or putting it on a more visible locatoin of a web page or adding advertisements/banners and then analyze changes to the user behavior.

Although the visual images represent preliminary results from our fusion of mining and visualization, we have endeavoured to use suitable guidelines and heuristics based on perceptual, cognitive and aesthetic criteria, for example using psychological notions of hot or cold color to guide assignment of color to nodes and edges having high/low numbers of hits, respectively. Using different diagrams we can focus on each of three link classes we find interesting: hot nodes/edges, cold nodes/edges, and edges connecting two different types of clusters i.e. hot clusters with high number of hits as opposed to cold clusters with low number of hits. This is valuable for web masters to make decisions and analyze changes of dynamics by informed or exploratory addition of edges between these two clusters. In future work we intend to use techniques such as splatting, spot noise, and/or filtering edges/nodes which are less important, in order to refine and clarify our understanding. As Hamming (1973) said,"The purpose of computation is insight not number". Likewise"The purpose of visualization is insight, not pictures" [2]. A detailed assessment of the utility of our diagrams is beyond the scope of this paper and we leave it for our future work. Recent studies lends support to our approach of discriminating different types of cluster structures [27].

5 Future Work

There is considerable room for improving the visualization of these results. First, at the algorithmic level, the scale and complexity of the graphs produced from the data mining stage still have the potential to embarrass available graph layout algorithms. There is often a tension in the design of algorithms between accommodating a wide range of data, or customizing the algorithm to capitalize on known constraints or regularities, and in the case of web log data, knowing more about the kind of graph that is to be drawn may help in simplifying the layout process. On the human side, further thought is needed on the mapping from data attributes to visual attributes, in particular where the visualization is superimposing access properties above the basic site structure. Part of this work can and should be based on known characteristics of perception and principles of visualization design, however, the ultimate utility of the representation will only become apparent once it is assessed through controlled experiments, and this will require time and a more polished version of the user interface.

A number of further tasks have already been mentioned in the text, the following could be added:

- A perceptual and logical appraisal of the visualizations relative to better understanding of specific user tasks.
- (Empirical) usability tests of the visualizations.
- Demonstrating the utility of web mining can be done by making exploratory changes to web sites e.g. adding links from hot parts of web site to cold parts and then extacting, visualizing and interpreting changes in access patterns. This may also require running our implementation on logs obtained over longer period of time.
- Visualizing output of related systems e.g. navigation predictors, recommender systems, browsing simulators, user modeling/profiling. Likewise a broad range

of AI algorithems, learners, probabilstic algorithms and more can be visually correlated with real user behavior in an unintrusive manner. Visualization of new Relational Probabilistic Models is of particular interest because of inclusion of relational structure of web sites in our techniques.
– Visualization of Link analysis and comparison of PageRank and Hub/Authority against actual web usage in line with Web Structure Mining. Also Web Content Mining can be introduced to our architecture in parralel to WWWPal suite.
– We remove web bot request in early stages of pipline, one interesting application is to visualize access path of those.

6 Acknowledgement

We would like to thank Prof. Mukkai S. Krishnamoorthy for his help on `WWWPal` [32].

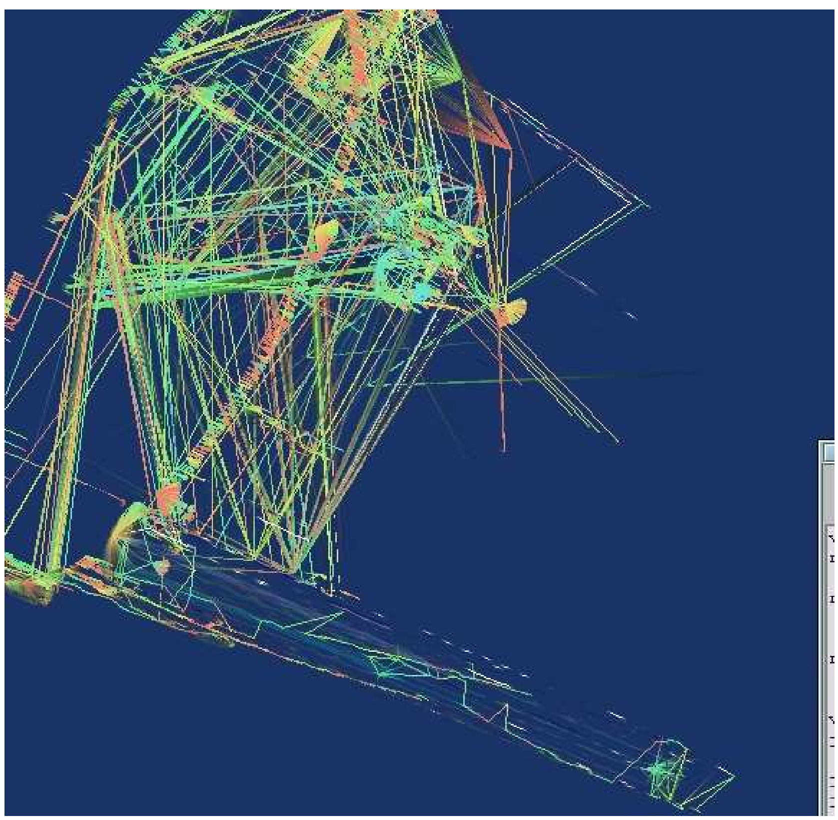

Fig. 2. Phenomena: Drilling Down, Cone and Funnel. User's browsing access pattern is amplied by coloring at the bottom of the picture.

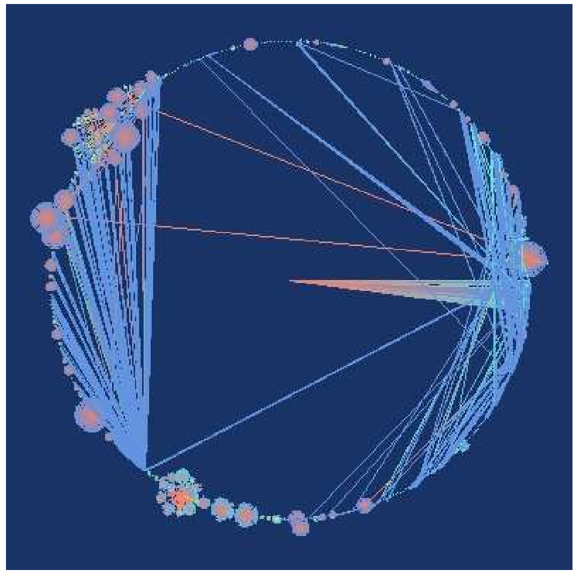

Fig. 3. 2D visualization with Strahler Coloring.

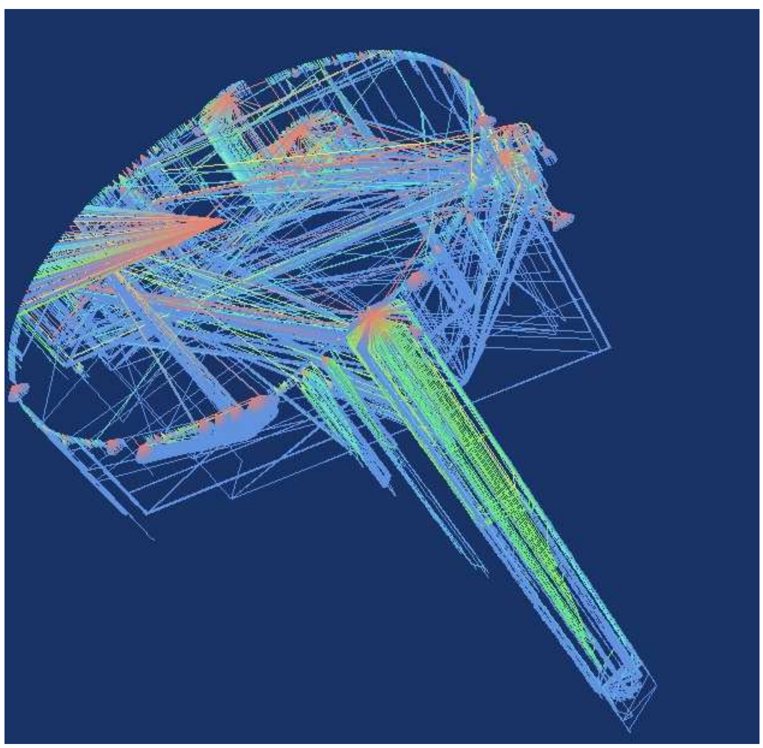

Fig. 4. Interactive 3D visualization of Drilling Down: Circular basement of the cylender like result is similar to figure 3, adding third dimension enables us visualize more information and clarifies user behavior in and between clusters.

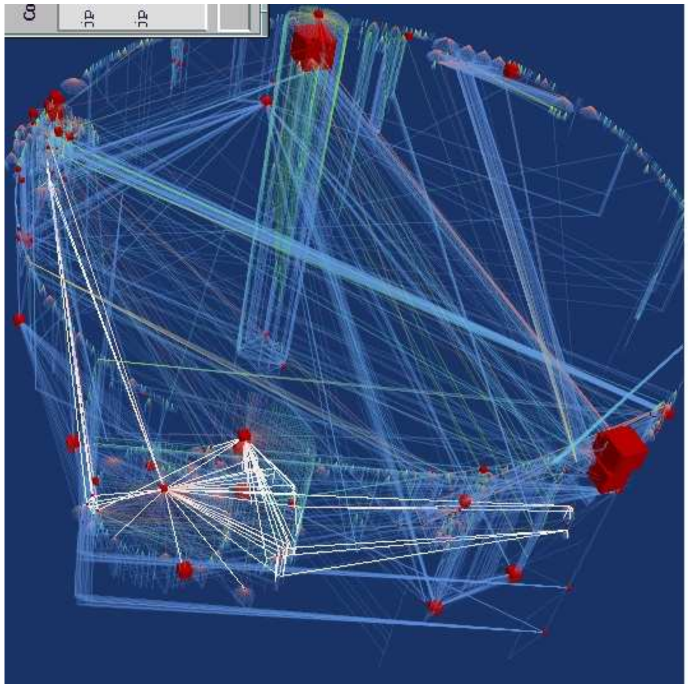

Fig. 5. Visualization of Mined Frequent Trees/Sequences: White lines are results of data mining algotithms i.e. frequent patterns.

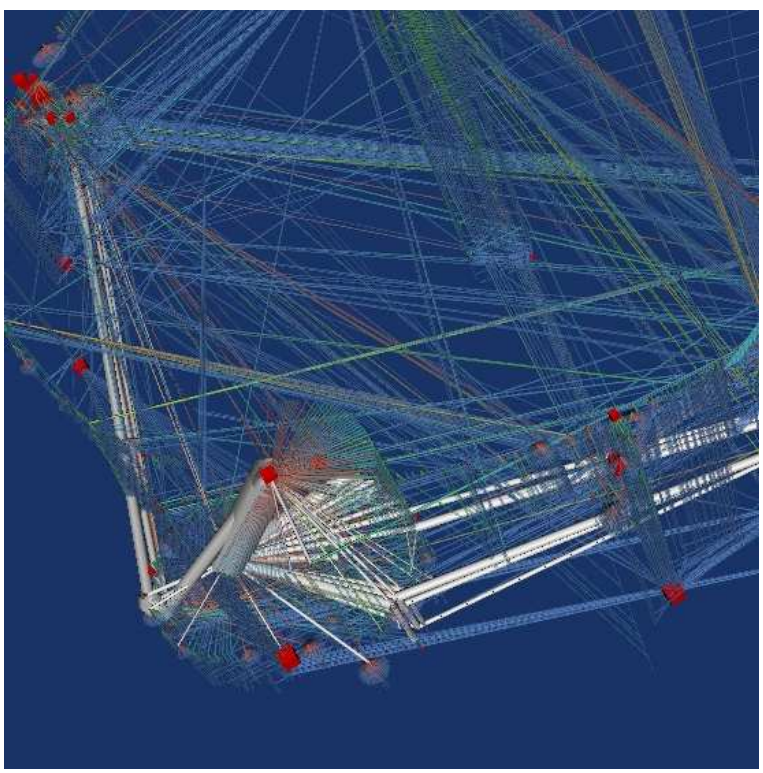

Fig. 6. Frequent Patterns visualized as Stream Tubes: Using tickness of Stream Tubes, we can add another dimension i.e. frequency of patterns to figure 5.

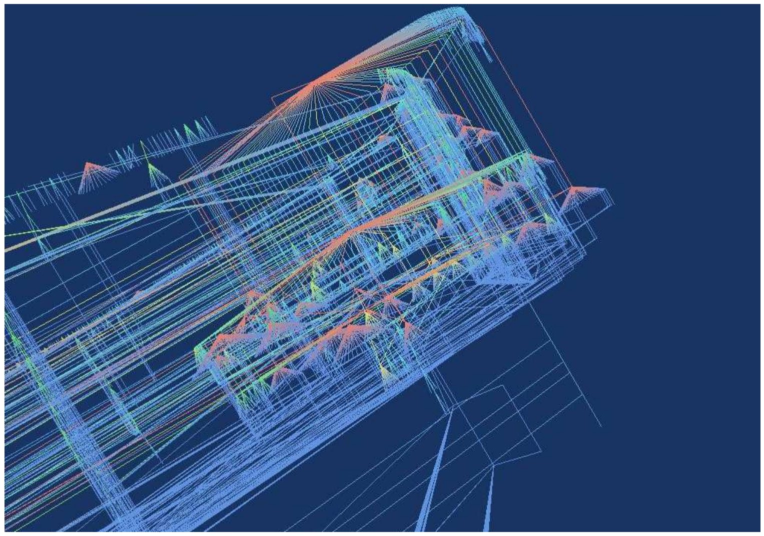

Fig. 7. Cone Trees.

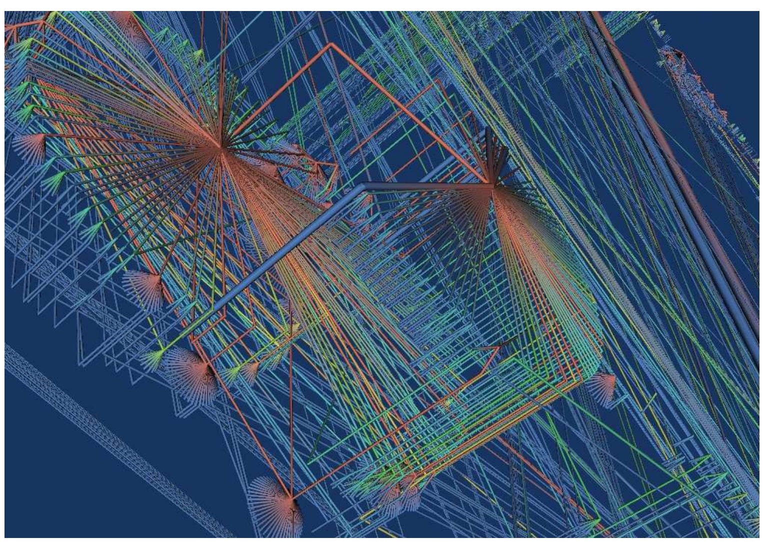

Fig. 8. Using click stream tubes on a zoom view.

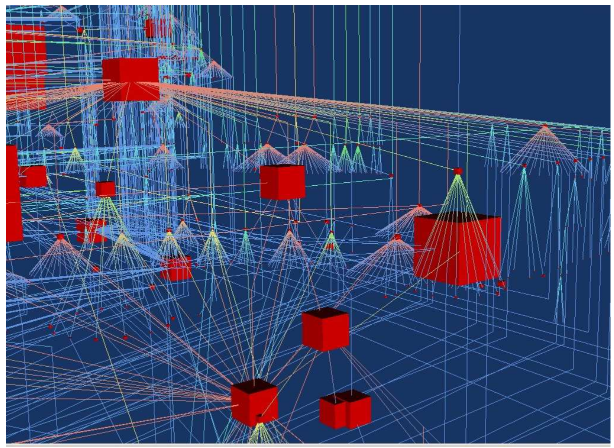

Fig. 9. Using Glyphs: Cube shaped glyphs with proportional size are put on nodes for easier selection by mouse.

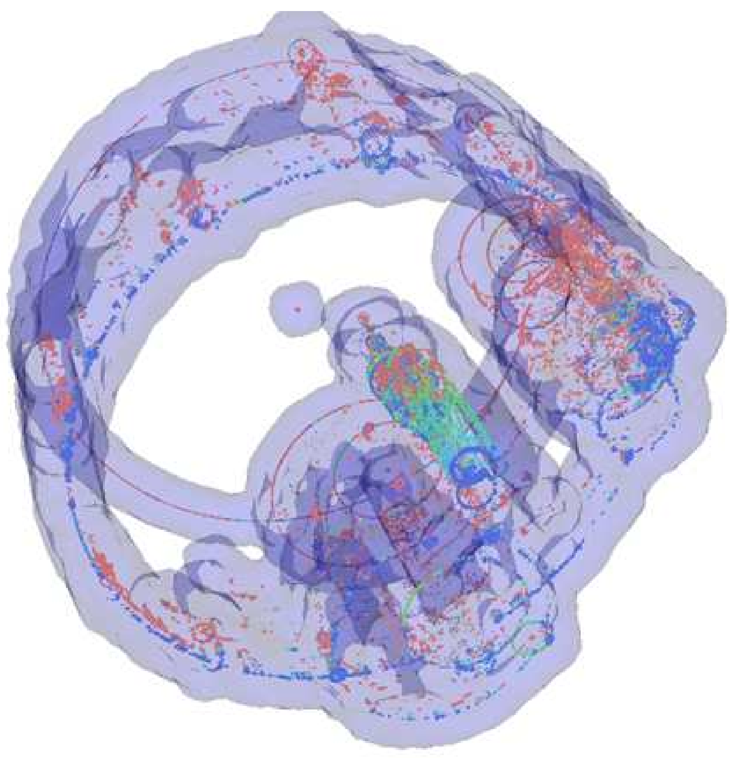

Fig. 10. Web Usage Superimposed on Web Graph with about 400,000 nodes. Semi-Static Web Graph extracted by webbot makes a gray basement for colored frequent paths extracted from Web Logs using Web Mining Algorithms. Clusters with/without frequent access are easily identified

References

1. David Auber, Using Strahler Numbers for Real Time Visual Exploration of Huge Graphs, International Journal of Applied Mathematics and Computer Science,2002.
2. S.Card, J.Mackinlay, and B.Shneiderman, Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann, San Francisco, 1999.
3. D.J. Duke, Modular Techniques in Information Visualization, Proceedings of the 1st Australian Symposium on Information Visualization, Volume 9 of Conferences in Research and Practice in Information Visualization, P. Eades and T. Pattison (Eds), pp. 11-18, Australian Computer Society, 2001.
4. T. Munzner, Drawing Large Graphs with H3Viewer and SiteManager, in Proceedings of Graph Drawing'98, 1998.
5. Daniel A. Keim, Information Visualization and Visual Data Mining. IEEE Transactions on Visualization and Computer Graphics 8(1), 2002
6. C. Shahabi, A. M. Zarkesh, J. Abidi and V. Shah, Knowledge discovery from user's web-page navigation, Proc. Seventh IEEE Intl. Workshop on Research Issues in Data Engineering (RIDE), pp. 20-29, 1997.
7. B. A. Huberman, P. L.T. Pirolli, J. E. Pitkow, and R. M. Lukose. Strong regularities in World Wide Web surfing. Science, 280:96-97, April 3, 1997.
8. G.G. Robertson, J.D. Mackinlay, S.K. Card. Cone Trees: Animated 3D visualizations of hierarchical information. Proc. of ACM SIGCHI, 1991.
9. E. Tufte, Envisioning Information. Graphics Press, 1990.
10. Alan Keahey, Stephen G. Eick: Visual Path Analysis. INFOVIS, 2002.
11. I. V. Cadez, D. Heckerman, C. Meek, P. Smyth, S. White. Visualization of Navigation Patterns on a Web Site Using Model Based Clustering, SIGKDD, 2000.
12. Kirsten Risden, Mary P. Czerwinski, Tamara Munzner, and Daniel B. Cook. An initial examination of ease of use for 2d and 3d information visualizations of web content. International Journal of Human Computer Studies, 53(5):695-714, 2000.
13. John Cugini,Jean Scholtz. VISVIP: 3D Visualization of Paths through Web Sites, WebVis, 1999.
14. http://www.nist.gov/webmet/
15. Tamara Munzner, Drawing Large Graphs with H3Viewer and Site Manager, Proceedings of Graph Drawing, 1998.
16. E. H. Chi, J. Pitkow, J. Mackinlay, P. Pirolli, R. Gossweiler and S. K. Card (1998). Visualizing the Evolution of Web Ecologies. ACM Conference on Human Factors in Software (SIGCHI), 1998.
17. Ed H. Chi, Stuart K. Card. Sensemaking of Evolving Web Sites Using Visualization Spreadsheets. INFOVIS 1999.
18. Paul Mutton, Peter Rodgers, Spring Embedder Preprocessing for WWW Visualization, 2002.
19. Linda Tauscher, Saul Greenberg, Revisitation Patterns in World Wide Web Navigation (1997). Proceedings of the Conference on Human Factors in Computing Systems CHI'97
20. Jim Sterne, Invited Talk: WebKDD in the Business World, WebKDD workshop of SIGKDD 2003.
21. F. H. Post, B. Vrolijk, H. Hauser, R. S. Laramee, and H. Doleisch, "The state of the art in flow visualisation: Feature extraction and tracking," Computer Graphics Forum, vol. 22, no. 4, 2003.
22. F. Reinders, I. A. Sadarjoen, B. Vrolijk, and F. H. Post, "Vortex tracking and visualisation in a flow past a tapered cylinder," Computer Graphics Forum, vol. 21, pp. 675–682, Nov. 2002.

23. F. Reinders, F. Post, and H.J.W.Spoelder, Visualization of Time-Dependent Data using Feature Tracking and Event Detection, The Visual Computer, vol. 17(1), pp. 55-71, February 2001.
24. F. Post, W. de Leeuw, I. Sadarjoen, F. Reinders, and T. van Walsum, "Global, Geometric, and Feature-Based Techniques for Vector Field Visualization," Future Generation Computer Systems, vol. 15, February 1999.
25. Tamara Munzner, Guest Editor's Introduction, IEEE Computer Graphics and Applications Special Issue on Information Visualization, 22(1), Jan/Feb 2002.
26. I. Herman and G. Melanon and M.S. Marshall, Graph Visualization and Navigation in Information Visualization: A Survey, IEEE Transactions on Visualization and Computer Graphics, 6(1), 2000
27. Miki Nakagawa, Bamshad Mobasher, A Hybrid Web Personalization Model Based on Site Connectivity, WebKDD workshop of SIGKDD, 2003.
28. B. Johnson and B. Schneiderman, Tree-maps: a space-filling approach to the visualization of hierachical information structures, Proceedings of Information Visualization'91, pp 284-191, IEEE.
29. W. Schroeder, K. Martin and B. Lorensen, The Visualization Toolkit: An Object-Oriented Approach to 3D Graphics, Prentice Hall, 1998. Also http://www.vtk.org.
30. John Punin, Mukkai Krishnamoorthy, Mohammed J. Zaki, LOGML – Log Markup Language for Web Usage Mining , in WebKDD Workshop of SIGKDD 2001.
31. `http://www.cs.rpi.edu/~youssefi/research/vwm`).
32. John Punin, Mukkai Krishnamoorthy, WWWPal System- A System for Analysis and Synthesis of Web Pages, Proc. WebNet 98, 1998.
33. Ivan Herman, Maylis Delest, Guy Melanon, Tree visualisation and navigation clues for information visualisation, Computer Graphics Forum, 1998.
34. M. Spiliopoulou and C. Pohle, Data mining for measuring and improving the success of Web sites, Data Mining and Knowledge Discovery, 5:85-14, 2001.
35. Mohammed J. Zaki Efficiently Mining Trees in a Forest, SIGKDD, 2002.
36. Mohammed J. Zaki, SPADE: An Efficient Algorithm for Mining Frequent Sequences, Machine Learning Journal, 2001