



Visual Mining Large Web-based Hyperbolic Space Using Hidden Links

Ming Hao, Meichun Hsu, Umesh Dayal, Adrian Krug*
Software Technology Laboratory
HP Laboratories Palo Alto
HPL-1999-20
January, 1999

E-mail: [mhao,mhsu,dayal]@hpl.hp.com
akrug@hpugrca.grc.hp.com

visualization,
hyperbolic space,
data mining

Many real-world KDD (Knowledge Discovery & Data Mining) applications involve navigation of large hierarchies, for example, internet resources, hot topics on the web, organization charts. Quite often users feel lost, confused, and overwhelmed with many complex displays. This paper discuss a new dynamic "hidden link" technique for linking a large highly connected graph in a simple hyperbolic space without cluttering the display. Only the primary directed non-cyclic links are shown. All other types of links (non-tree/cross-link) are hidden in the property of each node. These hidden links only appear when the user navigates/clicks on the node. "Hidden Link" allows a user to freely focus on the hierarchy of interest.

Internal Accession Date Only

*HP GmbH, Advanced Systems Lab, Ratingen, Germany

© Copyright Hewlett-Packard Company 1999

Visual Mining Large Web-based Hyperbolic Space Using Hidden Links

Ming C. Hao, Meichun Hsu, Umesh Dayal, Adrian Krug
(mhao, mhsu, dayal) @hpl.hp.com; akrug@hpugrca.grc.hp.com
Hewlett Packard Research Labs

Abstract

Many real-world KDD (Knowledge Discovery & Data Mining) applications involve navigation of large hierarchies, for example, internet resources, hot topics on the web, organization charts. Quite often users feel lost, confused, and overwhelmed with many complex displays. This paper discuss a new dynamic “hidden link” technique for linking a large highly connected graph in a simple hyperbolic space without cluttering the display. Only the primary directed non-cyclic links are shown. All other types of links (non-tree /cross-link) are hidden in the property of each node. These hidden links only appear when the user navigates/clicks on the node. “Hidden Link” allows a user to freely focus on the hierarchy of interest.

1. 0 Introduction

Recently, there has been a lot of research focused on visual mining of large hierarchies with millions of nodes, such as the World Wide Web structure, organization charts, and files systems, using a hyperbolic tree or space [1,2,3,6,7,8]. For example in Web, data mining, there is an immediate need for users to visualize the content and usage of the web. How to navigate through millions documents to access information on one display is a problem to solve. Hyperbolic space provides an elegant solution to display large hierarchies on a user screen. Hyperbolic space is different from the conventional approaches of laying trees on an Euclidean space. In Euclidean space, the area of a circle to contain nodes grows linearly. In hyperbolic spaces, the area of a circle grows exponentially with respect to its radius. As a result, The approach described in [2, 6] can handle a graph of over 20,000 documents on the web by using a focus and context scheme. Hyperbolic space allows a user to navigate through the nodes and to see the relationship of the visible portion of the space to the entire structure on one single display. It is not like MindMan [4], which requires multiple displays to represent large amount of data sets. The user has to click through display after display to find the information he/she needs.

In a directed hierarchical hyperbolic tree, nodes are organized along parent and child relationships. A parent can have many children. A child can have only one parent. For example, in an organization chart hyperbolic tree [1], each employee reports to one manager. A manager has many employees. However, in practice, this may not be sufficient; an employee may report to two managers (one regular manager, one temporary project manager). We need two paths: one to the regular manager, one to the temporary project manager.

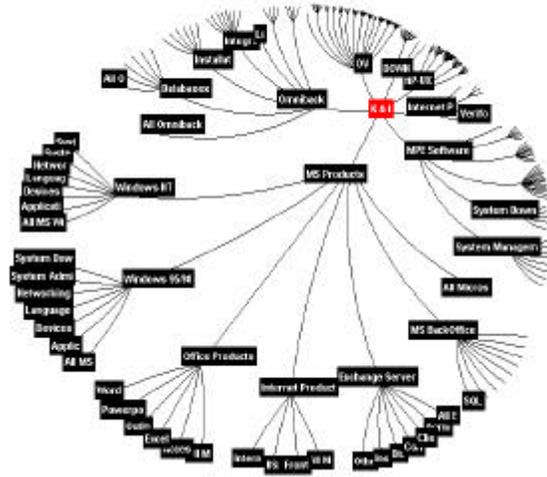
In a customer support web application, we use a hyperbolic tree [5] to organize large numbers of questions and answers in a hierarchical structure. Questions are parent nodes; Answers are arcs (pointers) to child nodes. A question can have several answers. An answer can lead to more questions and answers. However, it is sometimes necessary for an answer to link to another set of question/answer group (sub-space) which does not directly belong to the hierarchical hyperbolic tree.

Based on practical usage and design experience, we have discovered that strict hierarchical tree structures are too restrictive; often, we have a relationship that needs to be linked to a different branch of the tree. In the above examples, we need a separate path to represent the relationship between the employee with his temporary project managers, and to relate the answer with another set of questions. Commercial hyperbolic tree implementations (e.g. [5]) do not support such non-tree/cross edges. One solution would be to depict these relationships with additional lines. However, this solution could introduce thousands of lines and intersections. Also, for far away nodes that are off the screen, it would introduce “broken” lines (that do not terminate in a node on the current display). As a result, the hyperbolic space becomes very cluttered and difficult to visualize.

Figure 1 illustrates the difficulties of visualizing the multiple paths with existing methods.

Figure 1: Hyperbolic Space with & without Multiple Paths

1. Regular Tree

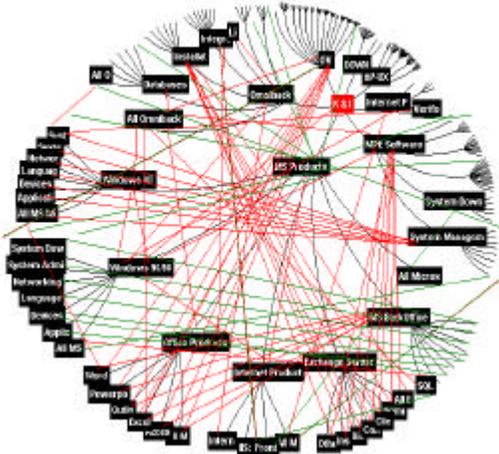


In a hyperbolic space, the far away nodes/edges (paths) are diminished when the user is not focusing on them. Those nodes will re-appear when the user warps the display to focus on or near them. The user can dynamically warp the display to focus on thousands of different nodes for navigation.

This graph is a non-cyclic hierarchical hyperbolic structure without multiple connected paths.

Contains: n nodes and $n-1$ edges

2. Multi-Paths Included



This graph is a hierarchical hyperbolic space with many highly connected paths.

There are many lines, intersections, and broken lines.

This hyperbolic graph becomes very cluttered

Contains: n nodes and $n-1$ primary edges
Plus many extra cross edges and intersections

2.0 A New Technique

How to visually mine a large web-based hyperbolic space with multiple paths while retaining the simplicity of the hyperbolic space is the problem we address. A new technique, called hidden link [9], has been developed. This

technique hides all the non-tree/cross paths in each node's property at the time the hyperbolic space is initialized. They become accessible and interactive only at the time of focus. The user can easily navigate through all possible paths

without tracing many lines and intersections. For example, when a user clicks on a child node that has a path – (e.g. link to his or her employee’s temporary manager), this technique automatically moves (maps and unmaps) the path to the temporary manager for the user to navigate through.

2.1 Definitions

In a directed non-cyclic hierarchical space, there is a primary graph, which links all the nodes in a tree form. These links are primary tree links. The others are non-tree/cross links in a highly connected graph. A node can have one incoming primary link and many non-tree/cross links. Figure 2 illustrates the following definitions.

Primary Path (tree-link) “AE”: A directed non-cyclic graphic link in a hierarchical hyperbolic space. Every node has exactly one primary parent (except for the root); the link from a node’s primary parent to the node is a primary path.

Secondary Path (non-tree/cross link) “AB”: A node may have additional secondary parents; the link from a node’s secondary parent to the node is a secondary path.

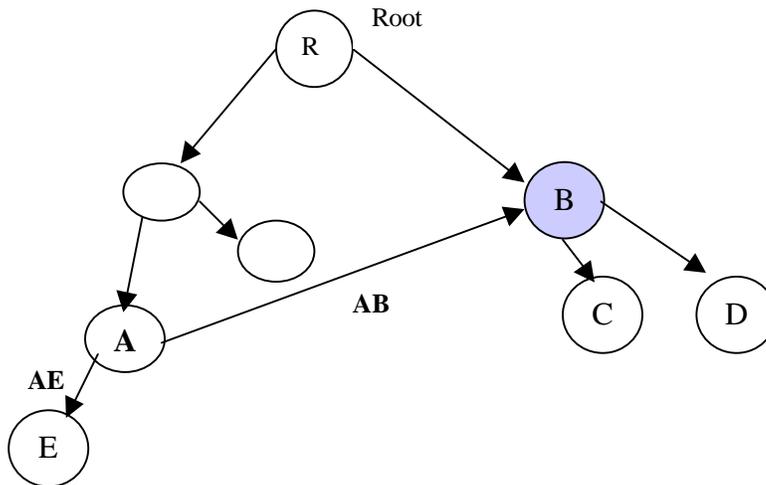
Hidden-Link Node: Node A contains both primary and secondary paths.

Primary Sub-Space Nodes: Defines sub-space parent node and its child nodes linked by a primary path. For example, node A and child node E linked by a primary path AE are the primary sub-space nodes.

Secondary Sub-Space Nodes: Defines sub-space parent node and its child nodes linked by a secondary path. For example, node B and its child nodes C and D linked by a secondary path AB are the secondary sub-space nodes.

As a result, a node (except for the root) has exactly one primary parent in a directed non-cyclic hierarchical hyperbolic space. If a node has more than one parent, then exactly one is designated as the primary parent, and the others are called secondary parents, the link from a node’s secondary parent to the node is called a secondary path. The path from a node’s primary parent to the node is called the primary path. For example, the path from the regular manager is called the primary path. The path from the employee to the temporary manager is called the secondary path. A node, which serves as a hidden link node, contains both the primary path and the secondary path

Figure 2: Hidden Link Definitions



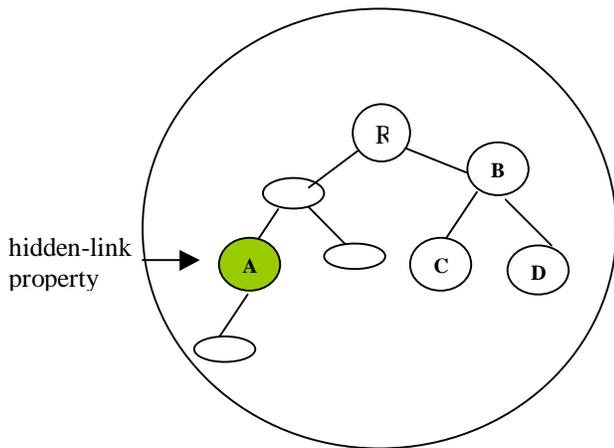
2.2 Hidden Link States and Processing Flow

Hidden link technique is built on a web-based client-server model with multi-threaded parallelism. It interacts with the user and hyperbolic space engines to allow instantaneous mapping and unmapping of secondary sub-space node for information visualization in a hyperbolic space structure.

This technique requires a hidden-link processor to manage the following five different states during the navigation of a secondary path. The hidden link property is defined at hyperbolic space initialization.

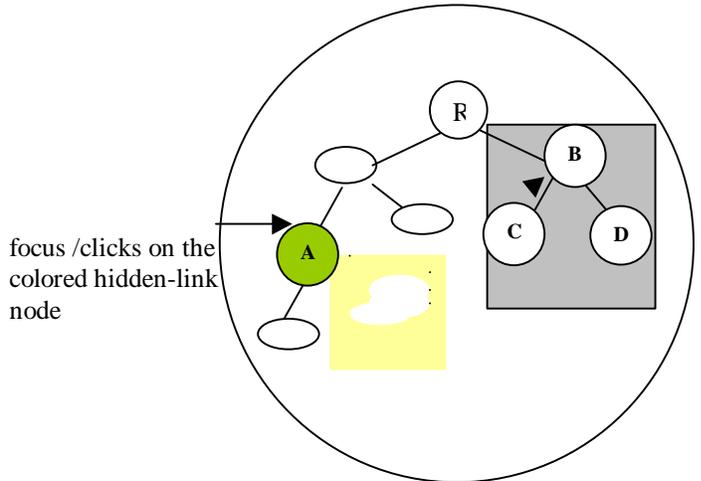
State 1: idle state

A hyperbolic space with hidden link has the same layout as it has without hidden links. There are no extra lines and intersections added in the graph. A hidden-link node is identified with a unique color.



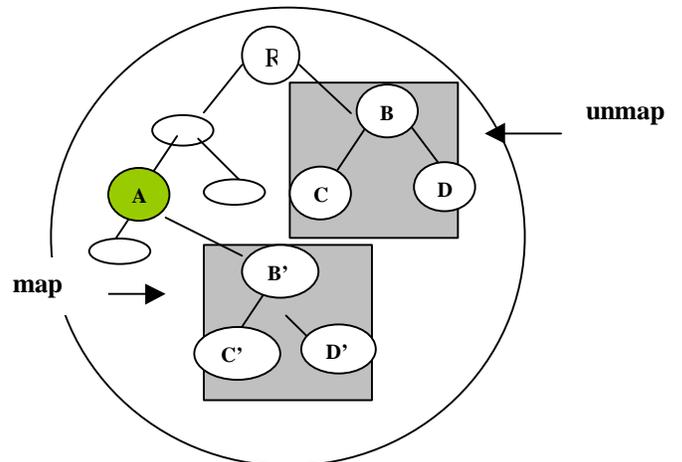
State 2: activate state

When the user is clicking on the hidden-link node, a number of temporary “secondary path” nodes will be drawn from that node. User can dynamically select a “secondary path” node for navigation.



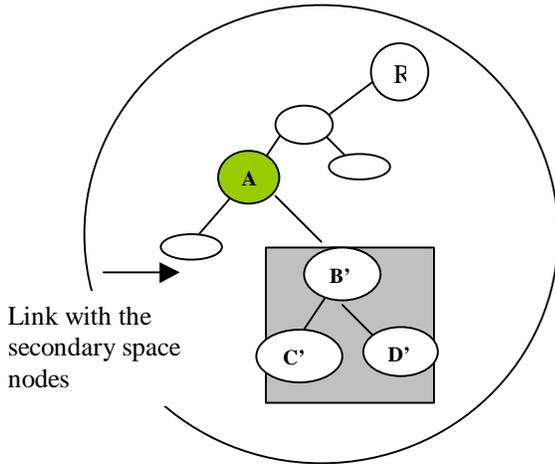
State 3: map/unmap (move) state

The hidden-link processor uses parallel mapping and unmapping methods to move the selected secondary sub-space with its children from their original structure to the hidden-link node.



State 4: Navigation State

After the selected secondary sub-space and its children are mapped under the hidden-link node, user can start his browsing and navigation.



State 5: Reset

At end of navigation, the hidden-link processor dynamically unmapps the secondary sub-space B', C' and D' from A. and maps primary sub-space B with its child nodes C and D back to it original structure. The hyperbolic space with secondary sub-space and its child nodes only existing during the secondary path processing. Checking of hidden-link node's ancestors to prevent cyclic map of the secondary path.

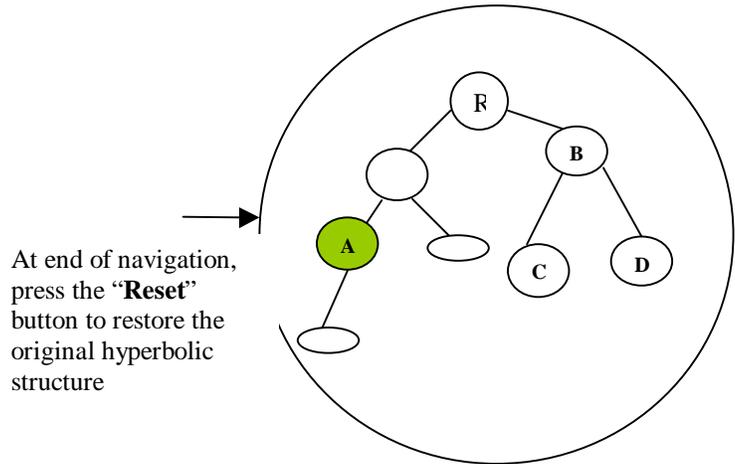
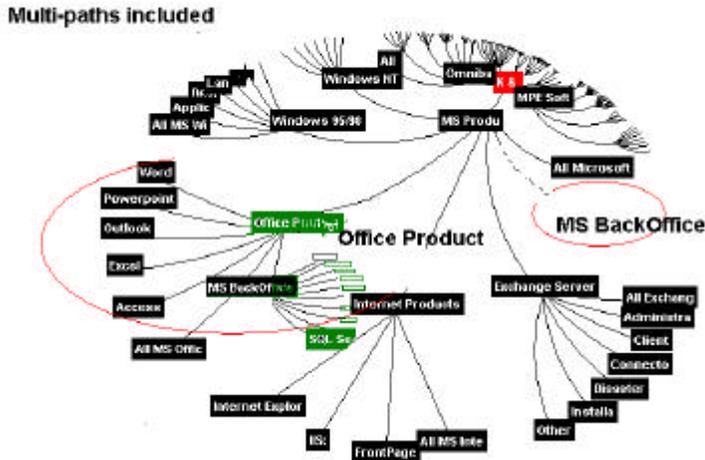


Figure 3: A New Multi-Paths Hyperbolic Space with Hidden Links



This multi-path hyperbolic space contains a hidden-link node “Office Product”. It has a secondary path to “MS BackOffice” (indicated as an empty circle)

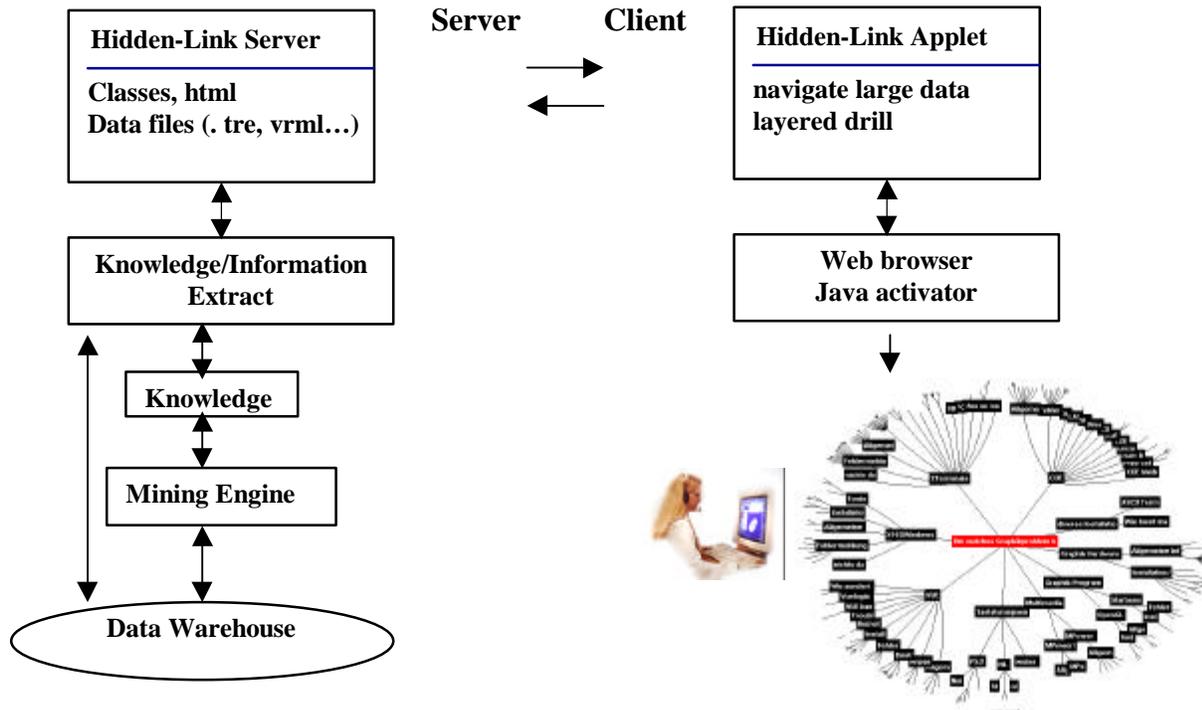
When a user is clicking on the “Office Product”, the “MS BackOffice” sub-space nodes will be dynamically moved from its original structure to the “Office Product” ready for mining.

Retain original hyperbolic simplicity without extra lines and intersections.

At end of navigation (reset):

- n nodes and n-1 edges
- no extra edges and intersections added.

Figure 4: “Hidden Link” Client-Server Web Structure



3.0 Web Data Mining Applications

Figure 4 describes the structure of a hidden link visualization server and a client on the Internet for data mining. We use a web browser with Java activator to dynamically create a large hyperbolic space on the web. The visualization Web interfaces are standard HTML and Java applet, which are used to explore relationships and to retrieve data within a region of interest. The server is integrated with the data warehouse and mining engine. The user is put in the driver seat at the client side to mine the knowledge results. It allows the user to dynamically access large hierarchies with complex links through HTML pages in a Web browser.

There are many data mining applications with large hierarchical information structures that can be mapped into hyperbolic spaces with multiple paths. We have cited several applications with

which we will use to illustrate these techniques (as shown in Figure 5).

3.1 Content and Usage Mining

The first example is to visualize the content and usage of the Internet [11, 12]. A hyperbolic space is constructed to present a topic hierarchy for over millions HEWLETT PACKARD UNIX documents linked to the web. The topic hierarchy is constructed by mining the content of the documents and session logs that record accesses to these documents. With hidden link capabilities, we are able to navigate a large highly connected topic hierarchy in a simple, uncluttered hyperbolic space on a web screen.

3.2 Customer Interview Web Service

The second example is a Hewlett Packard internal web based troubleshooting application called InterView [10]. We use a hyperbolic space

to organize and view large numbers of questions and answers in a hierarchical structure. Questions are parent nodes, while answers are arcs to child nodes. A question can have several answers. An answer can lead to another set of questions and answers. With the hidden link capabilities, users are able to navigate through answers to link to another set of question and answer group, which does not directly belong to the primary path of the hyperbolic space. The hidden link extension gives the hyperbolic space even the possibility to represents a general network and not only strict structured trees. User can easily follow the knowledge values to search for recent patches, technical tips, and editions.

In addition, statistical data can be mapped to the hyperbolic space using colored nodes or arcs. In this representation of data, like usage statistics, areas of new knowledge requests or even search result matches, author and users see an additional support to navigate through the knowledge.

4.0 Conclusion

Data mining applications in the 90's continuously face difficulties for visual mining of massive large highly connected data sets on

the Internet. To date, many practical applications have shown the usefulness of hyperbolic space [2, 3]. The hidden link technique described in this paper enables the mining of large hierarchies with multiple secondary paths.

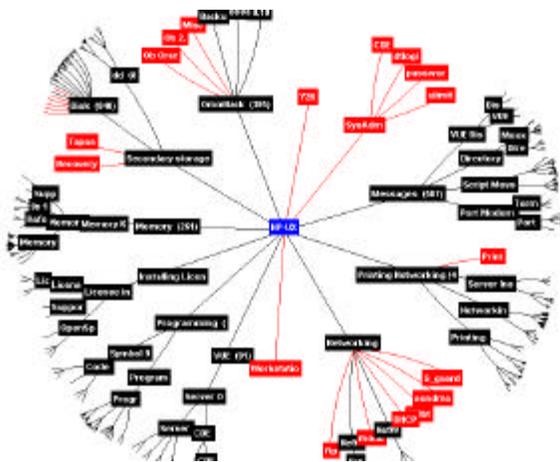
Hidden link enables the user to easily navigate through different links without being overwhelmed with large member of nodes and paths. The technique has been implemented in several prototypes at Hewlett Packard research laboratories.

5.0 Acknowledgements:

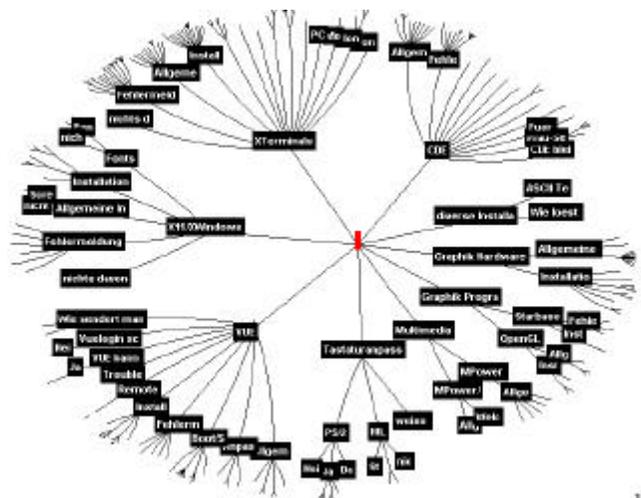
Thanks to Mary Loomis and Martin Griss from the Hewlett Packard Software Technology Laboratory for their suggestions and encouragement, Jeff Holmbeck, Hatold Shinsato from "Insight Software, INC" for technical discussions and for allowing us to use the hyperbolic tree toolkit for our experiments. Also thanks for Milon Mackey from the Hewlett Packard Computer System Laboratory for suggestions and reviewing.

Figure 5: Application Examples in Data Mining

(1) Topic & Content Hierarchies Sample



(2) InterView Application Sample



6.0 References:

[1] Charlie Gunn, "Discrete Groups and Visualization of Three-dimensional Manifolds" ACM 1993

[2] Tamara Munster, "Exploring Large Graphs in 3D Hyperbolic Space" IEEE Computer Graphics. Vol. 18, Number 4. 1998

[3] Site Manager 1.1 product information from Silicon Graphics. 1998.

[4] MindMan is a visualization and organizational tool to help individuals to organize, generate and learn ideas and information. 1998

[5] The Hyperbolic Tree Toolkit is a product from Inxight Software for exploiting large amounts of information. 1998

[6] John Lamping and Ramana Rao, "Laying out and Visualizing Large Trees Using a Hyperbolic Space". ACM /UIST'94, 1994

[7] George G. Roberson, Jack D. Mackinlay, and Stuart K. Card, "Cone Trees: Animated 3D Visualizations of Hierarchical Information", 1991 ACM.

[8] Joe C. Pinheiro, Don X Sun, "Methods for Linking and Mining Massive Heterogeneous Databases, KDD98

[9] Ming C. Hao, Meichun Hsu, Umesh Dayal, Adrian Krug, "A Technique for visualizing Large Web-based Hierarchical Hyperbolic Space with Multi-Paths." Invention disclosure pending, HEWLETT PACKARD Palo Alto Research Lab, 12/1998

[10] Adrian Krug: "InterView: Knowledge Content & Program Architecture" Hewlett Packard SW-Support Delivery Engineering, 1998

[11] Qiming Chen, Parvathi Chundi, Umesh Dayal, Mei Hsu, "Dynamic Agents for Dynamic Service Provisioning" Int. Conference. 1998

[12] Malu Castallanos, Qiming Chen, Parvathi Chundi, Umesh Dayal, Mei Hsu, Mike Lemon, Jim Stinger, "Component Advisor: A tool for automatically extracting electronic component

data from Web databsheets", W3C Workshop, 4/1998