

Information Visualization

Ganesh S. Oak (ganesh_oak@hotmail.com)

Fall 2001

*Project for CIS 732 (Design of Interactive Systems)
by [Prof. Turoff](#)*

Table of Contents

Abstract.....	1
Introduction.....	2
Data Types.....	4
One-Dimensional.....	4
Temporal.....	4
Two-Dimensional.....	4
Three-Dimensional.....	4
Multi-Dimensional.....	4
Tree.....	5
Network.....	5
Visualization Techniques.....	7
Two- and Three-Dimensional.....	7
Geometrically Transformed.....	7
Icon-based.....	7
Dense Pixel.....	7
Interaction and Distortion Techniques.....	8
Overview.....	8
Zoom.....	8
Filter.....	8
Details-on-demand.....	8
Relate.....	8
History.....	8
Extract.....	8
Examples.....	9
Fisheye.....	9
StarTree (Hyperbolic Tree).....	9
Spherical/Hemispherical Tree.....	10
Information Visualizer.....	11
Evaluation and Suitability.....	13
Data Characteristics.....	13
Visualization Characteristics.....	13
Task Characteristics.....	13
New Techniques.....	14
Storytelling.....	14
Alternatives to the 'desktop' metaphor.....	14
Scopeware.....	14
Task Gallery.....	14
Conclusion.....	16
References.....	17

Abstract

As we deal with more and more information in the digital era, it becomes increasingly difficult to cognitively process the whole matter in order to make sense out of it. Technological advances – particularly those related to computer speeds, graphics (display) resolution, and storage – have made it possible to evolve computer-based techniques to graphically represent huge amounts of information in a concise and understandable way.

The intent of this paper is to present an overview some of past research in the field of information visualization, with emphasis on certain popular techniques, and compare various techniques. I will also try to identify the promising techniques that need further research.

Introduction

Computers, and especially the World Wide Web have enabled us to access enormous amounts of information, which was not possible previously. Current estimates are that about 2 million terabytes of information is being produced per year worldwide [HMI], almost all of which is in digital form. The data are in various formats such as relational data, textual data, multimedia rich content etc. Different types of users have different needs to process data. Business users typically focus on data mining and data warehousing. They view data as a priceless asset that would help them identify spending patterns and product trends, which would in turn provide a competitive edge. Expert-system researchers and academicians are interested in knowledge networks. Individual users (consumers) deal with email, digital pictures, digital music etc. It is necessary to have capabilities to view, search, retrieve, and filter information so that a precise piece of it can be used at the time it is needed, with minimal effort exerted by the user. Exploring collections of information, whether on the Web or on one's local machine, becomes increasingly difficult as the volume and diversity increases.

“A picture is worth a thousand words.”

Human beings can perceive visual information better than other forms such as audio. Also graphical representations are easier to understand than textual representations. The eye, the mind, and the hand seem to work smoothly in a cohesive manner when using a graphic display. Graphic displays provide opportunities for showing different types of relationships by various means such as by proximity, by color encoding, or by containment. Information Visualization is based on this fact. Information visualization involves the use of computer-supported interactive visual representations of abstract data to aid the human cognitive process [PARC].

The hardware industry has experienced incredible advances with respect to the processor speeds, graphics (display) resolutions, memory and storage, and other technologies, in the recent past. Many software techniques can now be thought of in practical terms because of the availability of the supporting hardware.

These two factors – application demands, and technological advances – have fueled innovation in the information visualization field.

Research challenge in the field of information visualization is to invent new metaphors for visually representing information by understanding the analysis tasks that they need to support so that we can then develop ways to manipulate these metaphors to make sense of the information [EICK 2001]. The most important principle for the information visualization would be ‘Overview first, Zoom and filter, then Details on demand’ [SHNEIDERMAN]. First, the user identifies interesting patterns (overview), and then focuses on one or more of them (zoom and filter). To analyze the patterns, the user drills down to access details of the data (details on demand). Many researchers have been actively working in this field to extend this principle in various ways.

Various visualization techniques can be classified using three criteria: the data to be visualized, the technique itself, and the interaction and distortion method (see Figure 1 [KEIM 2001]). A visualization technique must support at least one data type and at least one interaction method to be a useful tool. The following sections discuss these criteria.

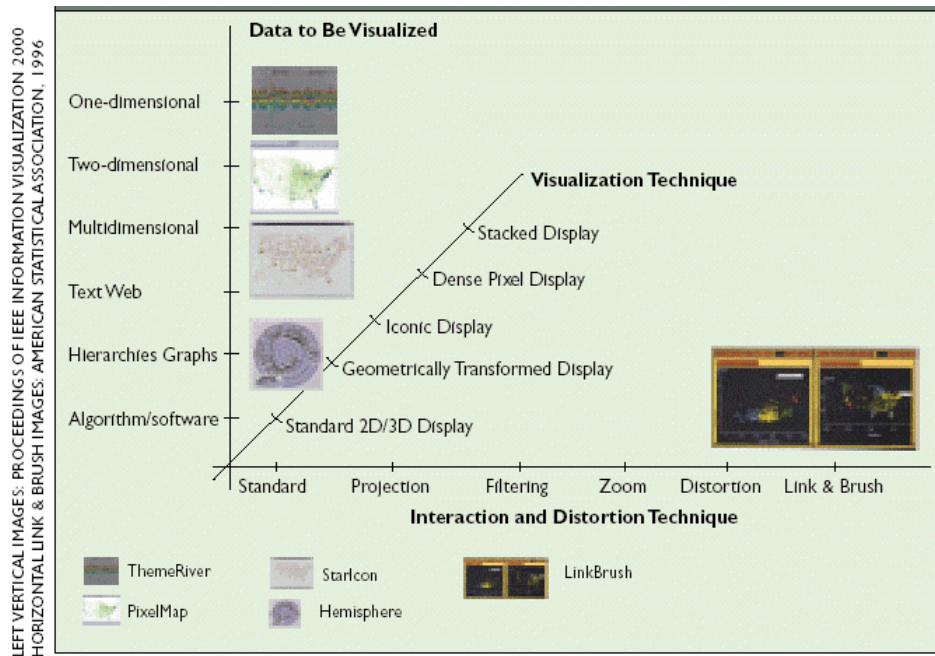


Figure 1 Classification of visual data exploration techniques [KEIM 2001]

Data Types

The information that we deal with can contain many different types of data. How we represent the information depends on the type of data that comprises the information. Shneiderman has described the prominent types that are typically used in computer applications [SHNEIDERMAN], [OLIVE]. Many different visualization techniques were developed for representing specific data types. These data types are described below (adopted from [SHNEIDERMAN] and [OLIVE]).

One-Dimensional

One-dimensional (i.e., linear) data is most commonly found in sequential lists, and textual documents. E.g., search results list, program source code file. Visualization challenges in this type of data are two-fold. One is that the data could be very large, requiring excessive scrolling to go through the entire collection. For long lists, it is necessary to have an overview of the entire document, with a capability to zoom in on a particular position. The other challenge is that each data item could lead to more detailed information. The degree of details (e.g., number of attributes) for each element could necessitate different views of the same content.

Temporal

Temporal data has time as the one dimension. Time lines are used very widely and are critical in many cases such as medical records and real-time systems. Hence they are treated separate from one-dimensional data. A typical example in the IS/IT industry is project management tools.

Two-Dimensional

Two-dimensional (i.e., map) data is most commonly found in geographic information systems. Two-dimensional data types should not be confused with two-dimensional representation. Since the conventional computer screens and paper are two-dimensional, two-dimensional representation is the most commonly used technique for visualization.

Three-Dimensional

Three-dimensional (i.e., real world) data is the most common type familiar to us in the real world. All the real-world, physical, objects are three-dimensional. Hence, there is a high demand to represent such data as we attempt to incorporate computers in all facets of life. Computer-aided design systems (CAD) for engineers, architects are built to handle complex three-dimensional relationships. When viewing 3D data, the user must cope with their position and orientation and handle the problems of occlusion. Occlusion could be a limiting factor for the perceived usefulness of a visualization tool. VRML (Virtual Reality Modeling Language) is very popular in applications that deal with 3D data.

Multi-Dimensional

Relational- and statistical-database contents are manipulated as multi-dimensional data. There are many applications that use multi-dimensional data, so this area is the most demanding when it comes to visualization techniques. The parallel coordinates technique [INSELBERG 1990] is quite an innovative way of representing multi-dimensional data in two dimensions. The TableLens [TABLELENS] shows multi-dimensional spreadsheet data in graphical representation along with textual cues within a single

easy-to-understand view. Users can chose and manipulate regions of the table to get all the details of one data element. The starfield display technique developed by Shneiderman (now it is called spottfire) is quite popular for displaying multi-dimensional data [SPOTFIRE]. Figure 2 gives a snapshot of the FilmFinder software that is based on the starfield technique [SPOTFIRE].

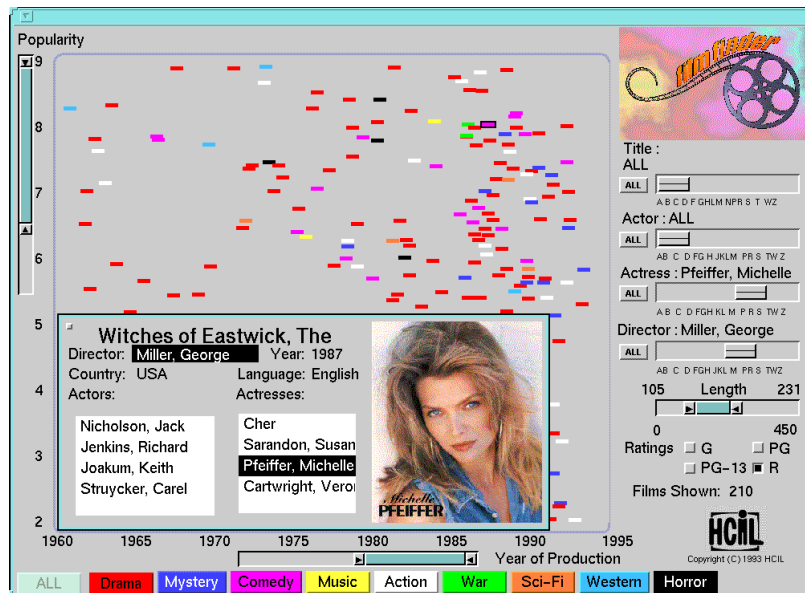


Figure 2 FilmFinder after selection of a single film [SPOTFIRE]

Tree

Hierarchical data structure is such that each element has a unique parent and zero-or-more siblings. The most logical way to arrange this data is in tree form. Hierarchical data is very diverse and is encountered in many forms. Organizational structure, and taxonomy are common examples of tree data. This data type can be directly related to the hierarchical index type [TUROFF]. Some of the popular techniques for visualizing tree data are cam-tree, cone-tree, hyperbolic-tree, and treemap. Figure 3 on next page gives an example of treemap that is used to represent a folder and its contents (including subfolders) on a Windows-based computer system. One well-known commercial implementation of treemap is the Market maps of SmartMoney (<http://www.smartmoney.com/maps>) (Copyright © SmaryMoney), which is famous in the financial industry for their slick Java-based (Java Applets) user interface.

Network

Data structures where one element is related to an arbitrary number of other elements are treated as network data. Finding a relationship among a set of data points in a network structure is not easy. To facilitate the relationships, structures could be categorized into different cases of networks (e.g., acyclic, lattices, rooted versus unrooted, and directed versus undirected). This enables researchers to develop algorithms to perform tasks on the specific structure categories such as finding the shortest or least costly path between two items or traversing the entire network. This data type can be directly related to the network index type [TUROFF].

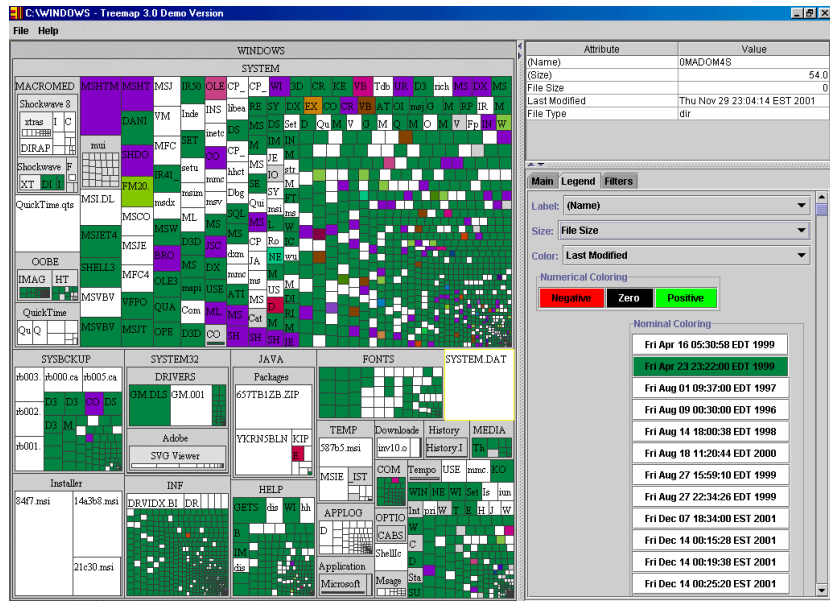


Figure 3 Treemap of a file folder on a Windows-based computer system

Visualization Techniques

The visualization techniques can be categorized based on how they arrange data and how they deal with multiple dimensions of data [KEIM 2001]. One technique could fit into one or more of the categories described below.

Two- and Three-Dimensional

These include standard 2D/3D techniques such as x-y plots, landscapes, etc.

In reality, all data visualization environments are displayed on a 2D surface. Two-dimensional representation is very effective in depicting direction, location, size, distance, etc. 2D techniques may use multiple layers to depict multi-dimensional data.

Volume visualization is currently the most widely used form of three-dimensional visualization. This involves rendering real world 3D objects into some form of computerized 3D representation. These objects could be viewed through immersive virtual reality equipment (e.g., dataglove), or be rendered on the conventional two-dimensional computer screen. Recently, 'artificial' ('synthetic') worlds have become popular. These are computer-enhanced virtual worlds that have the look and feel of a real 3D world, but do not truly exist in the real world. Activeworlds is an example of such a system [ACTIVEWORLDS].

Geometrically Transformed

This type of techniques uses geometric transformations and projections to produce meaningful visualizations. Techniques that deal with multi-dimensional data typically fall in this category. The parallel coordinates technique mentioned earlier is an example of this category.

Icon-based

Techniques that fall in this category display each data item as an icon, with dimension values rendered as features of the icons.

Dense Pixel

Techniques that fall in this category display each value as a color pixel and group the pixels belonging to each dimension into an adjacent area. By arranging and coloring the pixels in an appropriate way, the resulting visualization provides detailed information on local correlations, dependencies, etc. [KEIM 2001].

Interaction and Distortion Techniques

Data exploration, which is the ultimate goal of data visualization, also depends on interaction and distortion techniques [KEIM 2001]. Interaction techniques allow the user to interact directly with the visualization. These include Overview, Zoom, Filter, Details-on-demand, Relate, History, and Extract [SHNEIDERMAN]. Interactive distortion techniques support data exploration process by preserving an overview of the data during the drill-down (detailed view) operations. They purposefully distort the view to show more details of some portions of data while showing fewer details of other portions.

Overview

Overview strategies include bird's eye views that allow the user to see the entire collection, plus an adjoining detailed view that shows details of the selected item(s). Typically the overview contains a movable field-of-view area, which the user uses to control the contents of the detailed view.

Zoom

Users should be able to zoom in on a particular piece of information or a particular dimension of data. This would allow them to inspect particular details in depth.

Filter

It is necessary to be able to filter out uninteresting or otherwise unwanted items from the whole collection. This is especially very important in the context of Web, where possibility of information overload is very high.

Details-on-demand

Users should be able to get all the details of a particular item. An item should be identifiable after the collection has been cleared of any unwanted items.

Relate

Users should be able to identify relationships between different items in the collection. One of the most important applications of relationship identification is in data mining and data warehousing. These relationships help understand underlying patterns, from which users can extract business knowledge.

History

It is not uncommon to require a series of actions to arrive at a desired state in a system. This holds true for visualization systems also. Hence it is necessary to provide a history of user actions (menu selections, double-clicks, etc.) so that undo, redo could be supported.

Extract

Users should be able to extract data sets from the system. It is quite common to have only a few people involve in the data analysis, and have them communicate the results to others in the organization (their managers and up the hierarchy, peers, etc.). For easy communication, the system should support extracting data in commonly supported file formats (e.g., spreadsheets, charts, graphs, textual reports etc.).

Examples

There are many visualization tools available today. This section briefly demonstrates some of them.

Fisheye

A popular distortion technique is the fisheye technique. A fisheye camera lens is a wide-angle lens that magnifies nearby objects while shrinking distant objects. It is a valuable tool for seeing both 'local detail' and 'global context' simultaneously. Sarkar and Brown [SARKAR 1992] applied the fisheye approach to viewing and browsing graphs with a large number of nodes and vertices. Figure 4 shows a graph with 134 vertices and Figure 5 shows a fisheye view of the same graph [SARKAR 1992].

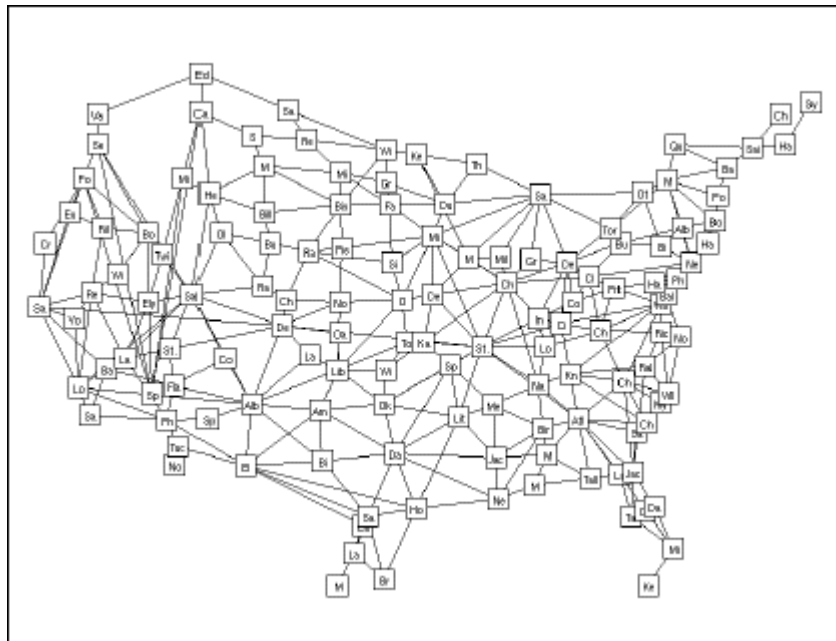


Figure 4 A graph with 134 nodes [SARKAR 1992]

StarTree (Hyperbolic Tree)

The hyperbolic tree browser is a variant of the fisheye (focus + context) technique that can be used for visualizing and manipulating large hierarchies. The tree is laid out on a hyperbolic plane and then this plane is mapped on to a circular display region. There is a natural correlation between the exponential growth in the number of nodes and the space available as the tree depth increases. Figure 6 shows an example of commercial implementation of hyperbolic tree [STARTREE].

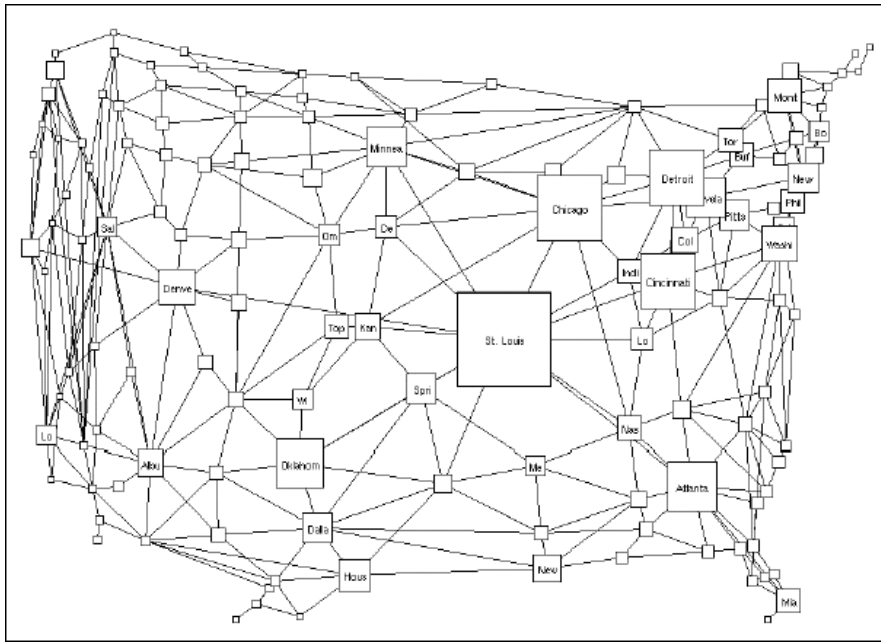


Figure 5 Fisheye view of the same graph [SARKAR 1992]

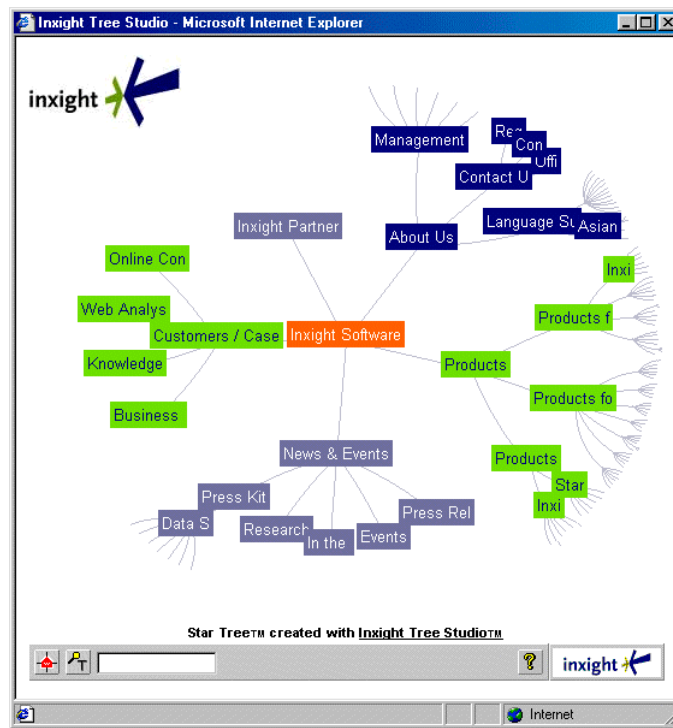


Figure 6 Star Tree view of the Web site <http://www.inxight.com>, provided by Inxight Software, Inc.

Spherical/Hemispherical Tree

This is similar to the hyperbolic tree. The difference is that a sphere or a hemisphere is used to map the data instead of a hyperbolic plane. This technique provides a nice

overview, good focus, even for very large graphs. Figure 7 shows an example [KEIM 2001].

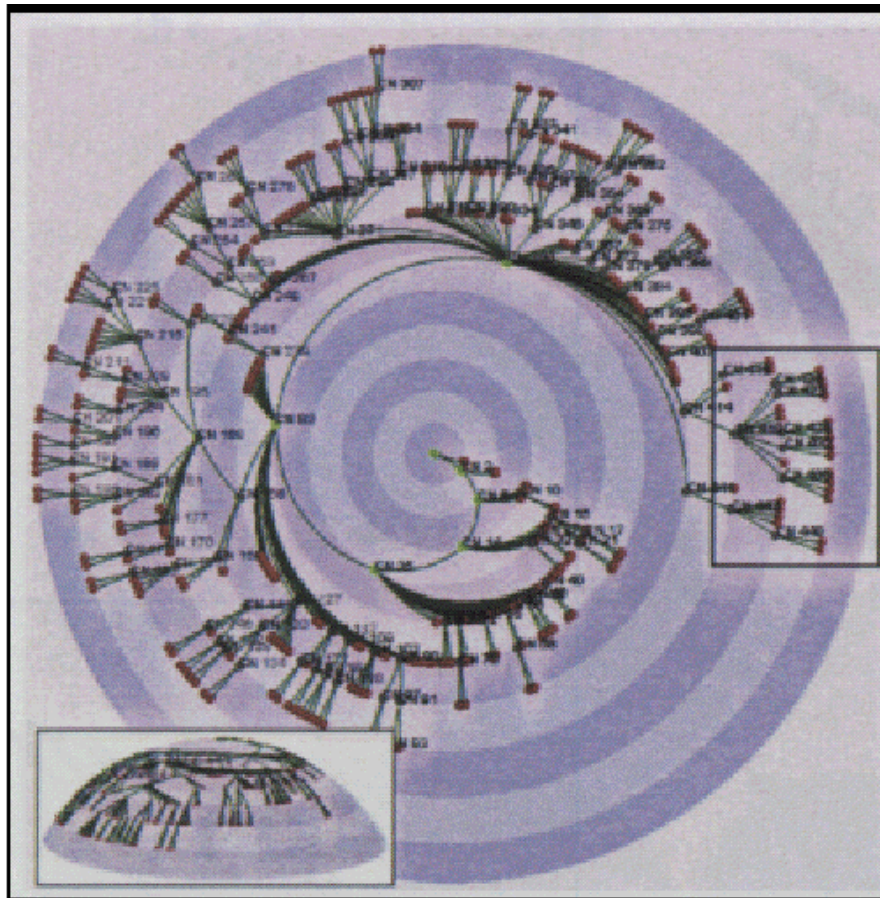


Figure 7 The hemisphere hierarchy visualization technique [KEIM 2001]

Information Visualizer

The Information Visualizer is an instantiation of a user interface for information retrieval systems. The system is based upon 1) information visualization, 2) 3D Rooms and 3) a Cognitive Coprocessor. Essentially, the information visualizer acts as an Information Workspace for information retrieval tasks [CARD 1991]. Figures 8 and 9 give examples of visualizations built into Information Visualizer [PARC].

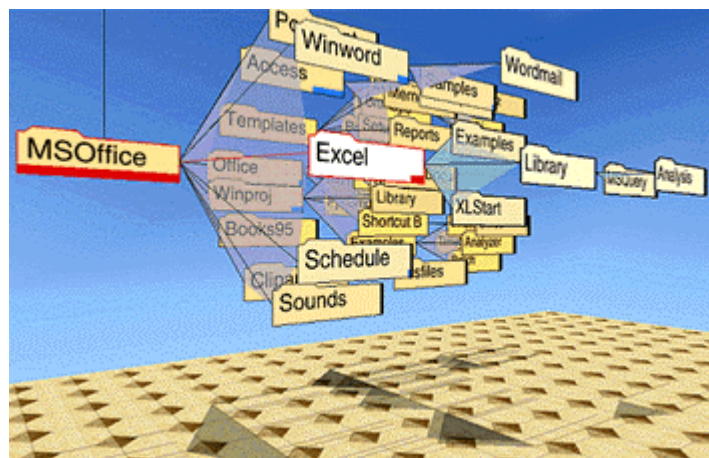


Figure 8 Cone Tree: A 3D interactive visualization of hierarchies [PARC]

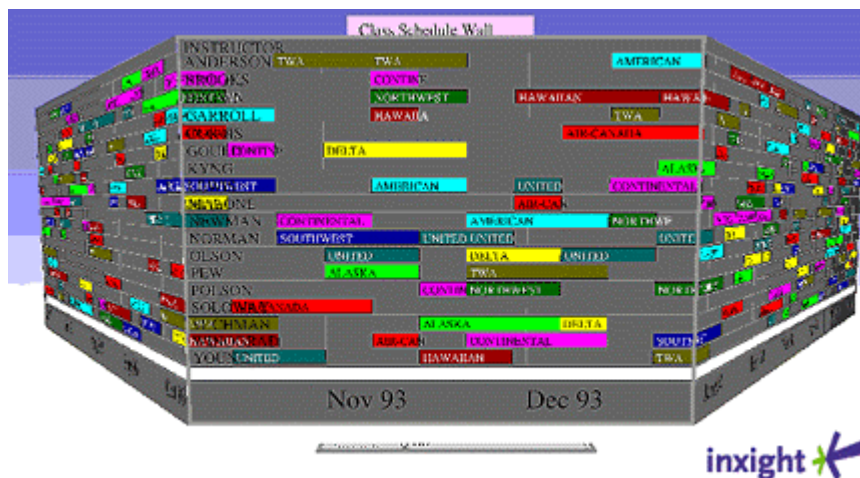


Figure 9 Perspective Walls: A 3D visualization of linear information [PARC]

Evaluation and Suitability

As is clear from above, there are numerous tools and techniques available for information visualization. This demands that various techniques and tools be evaluated to find their suitability for the system and task at hand. It is not possible to apply any technique without regard to certain constraining factors. Three types of factors need to be evaluated to find an appropriate visualization technique [KEIM 2001].

Data Characteristics

There are four characteristics of the underlying data that should be considered: type of data, number of dimensions of data, number of data items (volume), and category.

Different techniques are used for different data types. Some are specific for a given type, while others are generic in nature. E.g., hyperbolic tree would not be useful for representing one- or two-dimensional data. Similarly, all techniques are not suitable for any number of dimensions. E.g., icons cannot be used for a very large number of dimensions; parallel coordinates technique would not make much sense for data with less than three dimensions.

Visualization Characteristics

These include: visual overlap, and learning curve.

Visual overlap should be minimal. But, some cases might need more degree of independence than some other case where overlap is acceptable. The technique should have appropriate learning curve. If there is a large amount of multi-dimensional data, it might make sense to learn the more-difficult parallel coordinates technique. But if the data is not voluminous and the task is not repetitive, it might be advisable to spend a lot of time learning the technique.

Task Characteristics

The task characteristics that should be evaluated are: clustering, classification, and associations.

New Techniques

There are newer techniques evolving as researchers continue their work. This section describes such techniques. These techniques have not been widely used, as they are relatively new.

Storytelling

Storytelling is one of the most ancient forms of art. It is deeply rooted in the human culture, human philosophy, and human psychology [GERSHON 2001]. Technology has played an important role in storytelling. Various forms of entertainment such as press, radio, film, and television, make use of the continuously evolving technologies. As computer-related technologies advance in the field of rich multimedia, computers have earned a spot in the entertainment aspect of human life as well. Though it is widely accepted that pictures convey a great deal more than text, there are limitations on scenarios where pictures can be used. In certain cases, it is not practically possible to use pictures to convey a thought. In some cases, a well-told story can convey great quantities of information, for which even a series of pictures wouldn't suffice. This has led researchers to believe that storytelling can be effectively and efficiently used to aid visual presentation of information.

Alternatives to the 'desktop' metaphor

The traditional 'desktop' metaphor that we are so used to was initially developed to manage small data sets (Apple Computer's 1984 Macintosh, which had no hard-drive and its floppy drive capacity was 400 kilobytes). We have come to use the same metaphor for enormous amounts of information that we deal with in our everyday life today. This puts the burden on our brains to keep track of the huge collections of files, folders, and lists. To better utilize today's computing resources, researchers are toying with other concepts as a replacement of the 'desktop' [TRISTRAM 2001].

Scopeware

Fertig, et al developed Lifestreams [FERTIG 1996]. A lifestream is a time-ordered stream of documents. It is based on the 'diary' metaphor, in that it functions as a diary of one's electronic life. The tail of the stream contains documents from the past. Moving toward the present, the stream contains more recent documents. The system is now commercially available as Scopeware. Figure 10 shows a snapshot of how files are arranged in the system [SCOPEWARE].

Task Gallery

Microsoft is currently researching a metaphor that employs 3D space. It is called Task Gallery [TRISTRAM 2001]. The metaphor is that of multiple rooms in which documents are 'hung' on walls until they are needed. Simple mouse movements help move from one 'room' to another, and infinite 'rooms' can be added. Figure 11 presents a screen shot of the Task Gallery [TRISTRAM 2001].

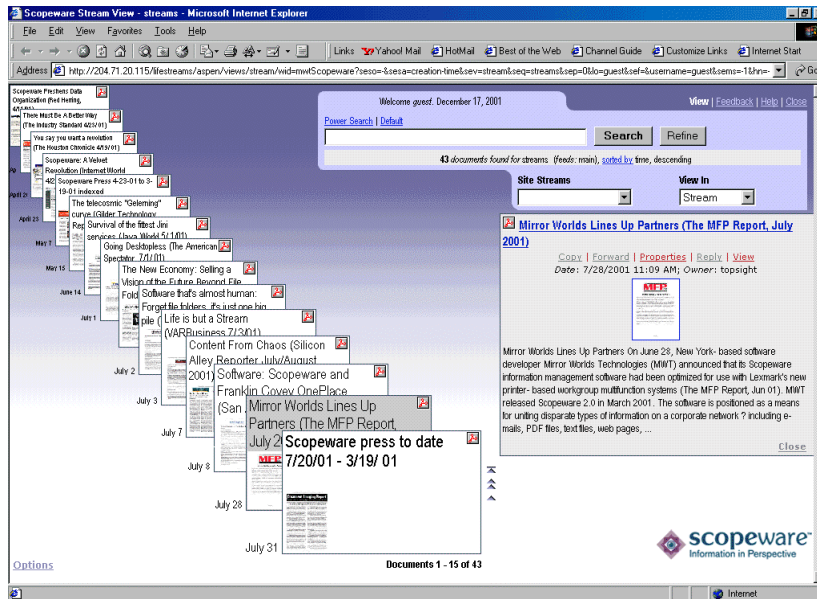


Figure 10 Files arranged in chronological order in the Scopeware system [SCOPEWARE]



Figure 11 Microsoft's new on-screen interface called Task Gallery: 3-D graphics help users remember where they placed their documents [TRISTRAM 2001]

Conclusion

Information visualization combines the aspects of graphics, human-computer interaction, and human-information interaction. Research challenge in the field of information visualization is to invent new metaphors for visually representing information by understanding the analysis tasks that they need to support so that we can then develop ways to manipulate these metaphors to make sense of the information.

It is important to evaluate and identify appropriate visualization tools for the task at hand. Applying any technique without proper study could lead to unsatisfied users and unfinished tasks.

Advanced information visualization techniques are still not widely used. Integration of these techniques with traditional techniques such as statistical-, simulation-, etc. techniques is the next step. This integration would allow us to better understand the information and use it in more effective way.

Integration of information visualization tools with different traditional applications is also very important. As we go forward to face challenges in the cyber world, we need to integrate visualization techniques with even greater number of systems such as security.

Old metaphors that are no longer valid should be eliminated or replaced with newer ones that are more apt for the current and future times.

All in all, there has been a lot of research in the field of information visualization, which, along with the technological advances, has opened doors to many new possibilities.

References

[ACTIVEWORLDS]	“A comprehensive platform for efficiently delivering real-time interactive 3D content over the Web”, http://www.activeworlds.com
[CARD 1991]	“The Information Visualizer: An information Workspace”, S. K. Card, G. G. Robertson, and J. D. Mackinlay, Conference on Human Factors in Computing Systems (CHI '91), ACM, 1991 pp. 181-188
[EICK 2001]	“Visualizing Online Activity”, Stephen G. Eick, Communications of the ACM, August 2001, Vol. 44, No. 8, pp. 45-50
[FERTIG 1996]	“Lifestreams: an alternative to the desktop metaphor”, Scott Fertig, Eric Freeman, and David Gelernter, Proceedings of the CHI '96 conference companion on Human factors in computing systems : common ground, April 1996, © authors
[GERSHON 2001]	“What Storytelling Can Do for Information Visualization”, Nahum Gershon, and Ward Page, Communications of the ACM, August 2001, Vol. 44, No. 8, pp.31-37
[HMI]	The “How Much Information?” project at School of Information Management and Systems, University of California, Berkeley, http://www.sims.berkeley.edu/research/projects/how-much-info/
[INSELBERG 1990]	“Parallel Coordinates: A tool for visualizing multidimensional geometry”, Alfred Inselberg, and B. Dimsdale, Proceedings of Visualization 1990, IEEE Press, pp. 361-370
[KEIM 2001]	“Visual Exploration of Large Data Sets”, Daniel A. Keim, Communications of the ACM, August 2001, Vol. 44, No. 8, pp. 38-44
[OLIVE]	On-line Library of Information Visualization Environments, http://www.otal.umd.edu/Olive/
[PARC]	Xerox PARC User Interface Research Group, http://www.parc.xerox.com/istl/projects/uir/default.html
[SARKAR 1992]	“Graphical Fisheye Views of Graphs”, Manojit Sarkar, and Marc H. Brown, Proceedings of Conference on Human factors in computing systems, ACM, 1992, pp. 83-91
[SCOPEWARE]	Mirror Worlds Technologies, Inc., http://www.scopeware.com , http://www.mirrorworlds.com
[SHNEIDERMAN]	“Designing the User Interface – Strategies for Effective Human-Computer Interaction”, Ben Shneiderman, Addison Wesley Longman ISBN 0-201-69497-2
[SPOTFIRE]	Ben Shneiderman, http://www.cs.umd.edu/hcil/spotfire/
[STARTREE]	Inxight Software, Inc.,

	http://www.inxight.com
[TABLELENS]	“A revolutionary technology that allows end users to interact with data directly”, http://www.tablelens.com , http://www.inxight.com
[TRISTRAM 2001]	“The Next Computer Interface”, Claire Tristram, MIT Technology Review, December 2001, pp. 52-59
[TUROFF]	“The Design And Evaluation Of Interactive Systems”, Murray Turoff, Working draft of the book

<End of document>