

[http://w.95898 Tm \(h\)TjDooDodg](http://w.95898 Tm (h)TjDooDodg)

INFORMATION MAPS: TOOLS FOR DOCUMENT EXPLORATION

Martin Dodge

Information Maps: Tools for Document Exploration



Introduction

So much has already been written about everything that you can't find out anything about it.

James Thurber (1961)

This white paper defines information mapping and reviews a range of examples. The focus is on two-dimensional interactive information maps that can be used to summarise large volumes of textual information, provide interfaces to browse the whole corpus and retrieve particular documents of interest. Information maps are being developed to tackle the modern-day challenges of information overload – too much, too fast and too unmitigated ¹.

The average person sitting at a networked PC has access to a huge (and hugely growing) array of textual data that can be utilised as a vital information resource in helping to make decisions. The data are in the form of documents on their own hard disk, messages pouring in through email and IM, from databases on intranets and digital libraries on the Internet and, of course, there is the expanse of the World-Wide Web, encompassing several billion static pages and an unknown volume of dynamically generated content. All these data are available at a mouse-click but they are beyond easy comprehension in their original forms, and can more often hinder, rather than help, us in finding answers and making appropriate decisions. To take the case of the Web, we are all familiar with using current Internet search engines to find an answer to a basic query and being overwhelmed with a lengthy list of matching pages. Too often, one retreats from this type of information-seeking with a sense of defeat, having been buried by the sheer volume of potentially relevant data and yet not being able to find answers.

Information mapping is one of the most useful means of assisting information seeking by summarising and visualising the non-visual content, structures and interrelationships of thousands of documents. Crucially, the information map should

¹ *Data Smog: Surviving the Information Glut*

Information mapping can be considered part of the larger research field known as *information visualization* ⁵.

There is, of course, a long history of information presentation using charts, graphics and map-like displays⁶, but it is only really since the 1990s that interactive graphics have been possible on affordable desktop PCs, thus enabling dynamic visual representations of large volumes of information accessible to large numbers of people. Previously, the user of an information map typically had access only to a static hardcopy product where the data selection, processing and presentation were predetermined and fixed by the map author.

Perhaps the two most widely cited ‘classics’ of this type of static information map, from the era before interactive graphic computing, are the London Underground diagram (the ‘Tube map’) designed single-handedly by Harry Beck⁷ in the 1930s and John Snow’s map⁸ of the cholera outbreak around the Broad Street pump from 1854. Beck’s Tube map is an incredibly useful map that enables millions of people to successfully navigate London’s complex subway network, but it also provides a very powerful framework for understanding the overall shape and geography of the city above ground. (This can sometimes be problematic as the map has gross and variable distortions of geographic scale.) Many of the spatial design metaphors developed by Beck are still used today on network and subway maps around the world. Snow’s map illustrates the potential of mapping (in this cases on a geographic template) to reveal previously unknown patterns which provide genuine insight into the problem at hand, i.e. attributing the cause of cholera to infected water, as demonstrated by the cluster of outbreaks around a particular water pump.

⁵ *Readings in Information Visualization: Using Vision to Think*, edited by Stuart K. Card, Jock D. Mackinlay and Ben Shneiderman (Morgan Kaufmann Publishers, 1999).

⁶ See the exemplary work of Edward Tufte in reviewing and critiquing the field <<http://www.edwardtufte.com/>>.

⁷ The fascinating story behind the Tube map is told in



Dot map of cholera deaths

The three key advantages of well-designed information maps are:

We often have a vague notion of what we are looking for, and likely places to start looking and we can usually tell when we have found an answer. An approximate

entire set of data. In decision-making, the useful information is drawn from the overall relationships of the entire set.”¹⁰

Maps and cognitive models

The role of the map is to provide a tangible external representation that aids people in forming an internal mental model or cognitive map of a space¹¹. The map in this sense is simply a tool to aid understanding. The understanding actually happens in the brain. A good map can provide the vital ‘Ah-ha’ factor when one sees structure, patterns or find new connections that had not been apparent without the map. This is due in part to the fact that the human brain has excellent spatial processing capabilities (we navigate through a complex three-dimensional world without conscious effort) and powerful spatial memory (which enables us to remember where a multitude of things are, and how to find them again). As Professor Robert Spence notes,

“The relevance of spatial memory to visualization is clear. Since visualization is the creation and enhancement of an internal model, it is useful to know that spatiality as a basis appears to lead to a robust internal model, and therefore that phenomena having no inherent location could perhaps be assigned an artificial location – consistently used throughout – in order to enhance the likelihood of a robust internal model.”¹²

The cognitive processes at work in the acquisition of this insight from the map are still not fully understood by scientist, being the subject of ongoing research in cognitive psychology and behavioural geography. However, insights clearly do arise from mapping information, as many of us can testify from personal experience.

¹⁰ *Graphics and Graphic Information Processing*, by Jacques Bertin (trans. by Walter de Gruyter, 1981), page 64.

¹¹ *How maps work: representation, visualization, and design*, by Alan M. MacEachren (Guildford, 1995).

¹² *Information Visualization*, by Robert Spence (Addison-Wesley, 2001), page 60.

Map design

Information maps use a two dimensional display to represent data content, structures and relations via a range several spatial properties. The most important properties are:

- Area
- Position
- Proximity
- Scale

The *area* of graphic objects in the map is perhaps the most important representational tool as it is used to show the size or volume of different groups of data. The larger the area on the map, the more data it represents. The *position* of graphic objects can also have meaning - for example, they can be arranged into hierarchies, or ordered by time, size, importance/relevance or simply alphabetically. *Proximity* means that the closer together graphic objects are drawn on the map, the more alike they are. Distance on the map equates to some metric of similarity between documents. This gives rise to local neighbourhoods and regions in the map that are likely to contain documents with similar content and themes. The ability to create maps of a data-space at different *scales* is also vital as it allows different level of detail to be shown at appropriate times. Large-scale maps can provide a good overview by showing a large extent at low detail and then when required the user can use a smaller-scale map to reveal rich local details. Moving between scales can be by simply switching display or a smooth transition via some kind of a zooming function. The application of these spatial properties to summarise and represent information has been term *spatialization*¹³. Note, not all information maps make use of all these spatial properties.

Additional graphic properties such as colour, shape and labelling, for example, can encode further details of the data in the map. For example, colour is often used to show magnitude or change. Labels can be very useful in identification of specific items or regions of the map. And the shape of graphic objects in the display can

¹³ M 90016 Tc 0.0016 Tw Q vc36 -4.32 Td()Tj3rm79 Tw 29i901 W b -6.36 TdTc 0 Tw TDr7oav Td(is al 90016 Tc 0.00Tj-0.008dt

convey useful subsidiary attributes of documents. For example, the particular aspect ratios of rectangular tile symbols can encode data in terms of the width and height values. So, for example, long, skinny rectangles would be different to short, fat squares.

Reviewing Information Maps

The map is a help provided to the imagination through the eyes.

Henri Abraham Chatelain, *Atlas historique* (1705)

In the 1990s a range of different information maps has been developed. In this section of the white paper we review the significant examples:

Research prototypes:

- Treemaps (Human Computer Interaction Lab, University of Maryland)
- ET-Map (Artificial Intelligence Lab, University of Arizona)
- WEBSOM (Neural Networks Research Centre, Helsinki University of Technology)

Commercial products:

- ThemeScape (Cartia, Inc.)
- Map of the Market (SmartMoney.com)
- Visual Net (Antarct.ca Systems, Inc.)
- WebMap (WebMap Technologies, Inc.)

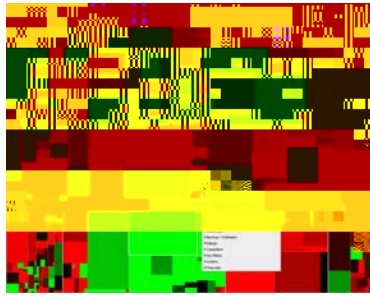
ThemeScape

Company: Cartia, Inc.

Address: Bellevue, Washington
State, USA.
No current website.

Status: Initial research at PNNL

Map of the Market



Company: SmartMoney.com

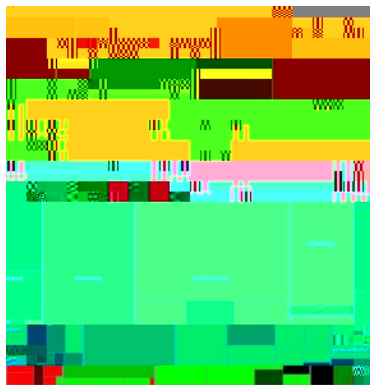
Address: 1755 Broadway, New York City, NY, USA..
www.smartmoney.com/maps/

Key developer:
Martin Wattenberg

Status: Launched at the end of 1998. Extended to include IPO and Mutual Fund maps as well stand alone Mapstation and Excel plug-in. Full product is online.

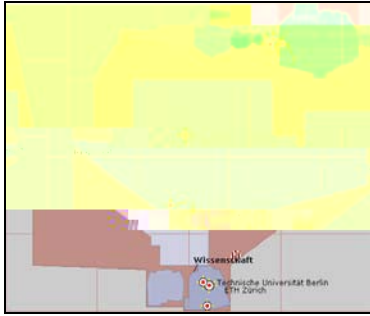
Further reading: *Visualizing The Stock Market*, by Martin Wattenberg. Proceedings of ACM CHI'99 Conference, 1999, pp. 188-189.

Show me the money: The Map of the Market, Mappa.Mundi Magazine, August 2001.



Map of the Market aims to answer a simple, but vital question for investor, 'how is the market doing today?'. It does this by mapping the performance of the stock market in a single visual snapshot showing the capitalization and changing stock price of over 500 publicly traded US companies. On one single map one can quickly gain a sense of the overall market conditions, yet still see many hundreds of individual data elements. This overall picture is very difficult to comprehend from more conventional listings of stock prices.

Visual Net



Company: Antarcti.ca Systems Inc.

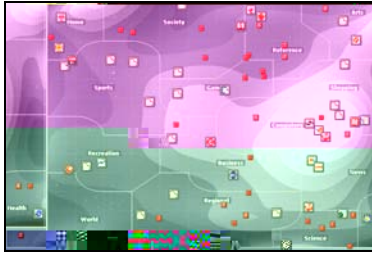
Address: 1198 Homer Street,
Vancouver, BC, Canada.
<http://antarcti.ca>

Status: Company formed in 1999 and Visual Net and Map.net launched in November 2000. Online demonstrations available mapping the ODP Web directory, the PubMed National Library of Medicine's database, and Canadian VC activity.

Key developers: Tim Bray, Dave Ashworth and Glen McMillan.

Further reading: [Visual Net:](#)

WebMap



Company: WebMap Technologies, Inc.

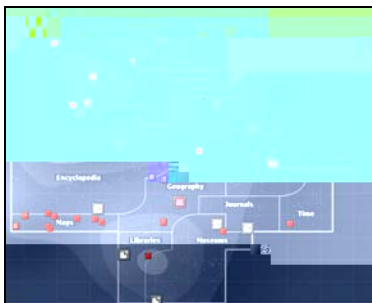
Address: 175 Portland Street,
Boston, MA, USA.
www.webmap.com

Status: Formed in November 1999 and initial map application launched in spring 2001. Development is ongoing focusing on travel, pharmaceutical and financial services. Free download of demonstration InternetMap.

Key developers: Michael Iron, Ohad Ranen and Roi Neustadt.

Further reading: [WebMap Technologies, Inc.](#) White Paper, April 2001.

Visualizing the Web, by Paul Heltzel. Technology Review, April 2001.

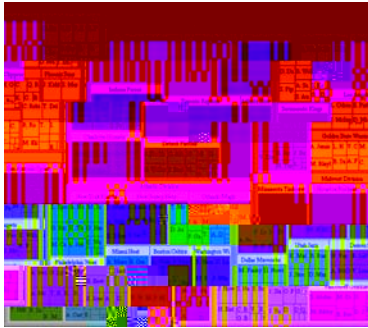


WebMap Technologies is developing an interactive, multi-level information mapping system. Their self-stated aim of their Information Mapping Technology is “*to make interaction with the Internet, intranets, extranets and wireless environments more useful and efficient through a visual rather textual format.*” The map comprises three layered components – categories, ratings and document icons. Their flagship mapping technology demonstrator, InternetMap, launched in the spring of 2001, maps more than 2.5 million websites classified by the Open Directory Project. The top-level map comprises the standard high-level categories. Through an interactive zoom you can explore more than 350,000 categories, as well undertake key word searches with the location of matching websites shown by small red icons. Your browser favorites are shown as icons on the map.

The system uses a well implemented client-side browser plug-in to deliver an interactive map of a collection of documents. Documents are grouped together into categories based on similarity of topic. These are represented on the map as irregular shaped regions. A terrain style map underlies the category regions and acts as a kind of ‘topography of popularity’. The higher the terrain (shown by the lighter shading contours) the more highly rated the documents are in that region. Various icon symbols are used to mark the location of particular documents, such as book marked documents or results from a conventional key word search. The map is multi-level and clicking on a region will smoothly zoom the view and set the display to this region and change scale to reveal more detailed categories.

The server side of the system can analyze documents, grouping them in categories based on content and positioning them using similarity proximity. The actual analytical techniques and clustering algorithms employed by WebMap are proprietary. The system also features a degree of map customization including the ability for the user to change categories of a document by dragging and dropping the icon and also a personalization engine that “

Treemaps



Research center: Human
Computer Interaction Lab (HCIL)

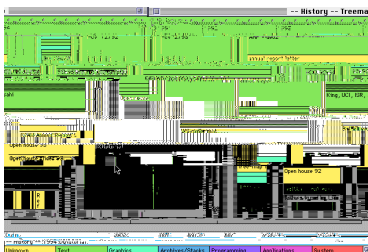
Address: University of Maryland,
College Park, MD, USA.
www.cs.umd.edu/hcil/

Status: Initial technique
developed by Shneiderman in
1990. Developed through 90s and
latest release of java-based
Treemap 3.0 is available for
commercial licensing. Research
into treemaps is ongoing at HCIL.
Free online demonstration is
available.

Key developers: Ben
Shneiderman, Brian Johnson and
David Turo.

Further reading: [Treemaps for
space-constrained visualization of
Hierarchies](#), by Ben
Shneiderman, November 2000.

[Treemaps: a space-filling
approach to the visualization of
hierarchical information
structures](#), by Brian Johnson &
Ben Shneiderman, HCIL Paper
HCIL-91-06, 1991.

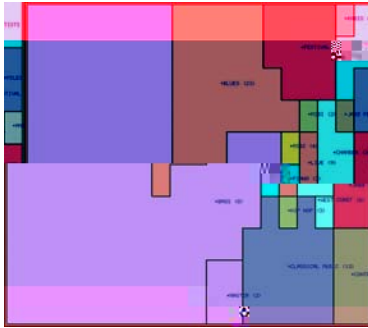


Treemaps are defined as a space-constrained visualization of hierarchies and they are one of the most well developed information mapping techniques having begun life back in 1990 and been refined and extended throughout the decade by researchers at HCIL. Treemaps show hierarchical data as a set of nested rectangular tiles that fill the whole display. Each data element is represented on the treemap as a single tile and key attributes of the data are represented by the size and colour of the tiles. Treemaps enable a large volume of information to be usefully displayed on a single map screen. In certain task treemaps have been shown to facilitate users in comparing data elements, gaining a sense of the whole data space and identify patterns and anomalies. The inherent hierarchical structure of the input dataset is represented in the treemap display by the nesting which is visually enhanced with border around each level in the hierarchy.

Shneiderman originally developed the treemap concept and algorithm in 1990 as a visual aid to the file management tasks on congested hard disk of shared PC. The hierarchical data was the directory structure of files. Each tile in the treemap display was an individual file on the disk, with the area of the tile proportional to the file size and the colour indicating different file types. Treemap therefore mapped the content of the disk and revealed which files (and directories) were taking up precious disk space. Shneiderman comments, *“I think treemaps are a convenient representation that has unmatched utility for certain tasks. The capacity to see tens of thousands of nodes in a fixed space and find large areas of duplicate directories is very powerful It does take some learning for novices to grasp the tree structure layout in treemaps, but the benefits are great.”*

Treemap algorithms have been refined to improve layouts and produce more legible displays. While a number of new implementations have been released throughout the 1990s. Treemaps have been applied to a diverse set of information domains, including financial data, basketball statistics, decision-making processes and network management.

ET-MAP



Research center: Artificial Intelligence Lab, University of Arizona.

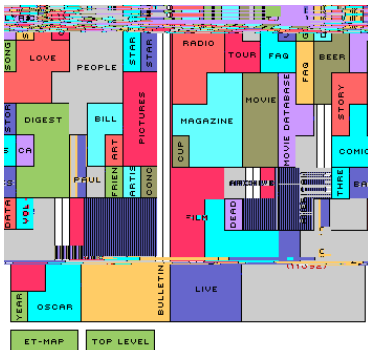
Address: Management Information Systems Department, University of Arizona Tucson, Arizona, USA.
<http://ai.bpa.arizona.edu>

Status: Research prototype developed in the mid 1990s. Development has ceased as efforts of the center are applied in other project in intelligent information classification and retrieval.

Key developer: Hsinchun Chen

Further reading: [Internet Categorization and Search: A Self-Organizing Approach](#), by Hsinchun Chen, Chris Schuffels, and Rich Orwig, 1996.

[A Map of Yahoo!](#), Mappa.Mundi Magazine, February 2000.



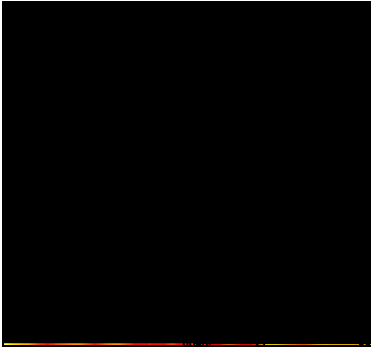
ET-Map is a prototype Internet homepage categorization system developed by the AI Lab. It is attempt at a meaningful multi-layered information map based on part of the Yahoo! directory.

They analyzed and automatically classified the content of 110,000+ entertainment-related homepages producing a multi-layered map with different subject regions. Larger subject regions occupy a bigger space on the map and conceptually related subjects are often grouped in close proximity. Regions which contain 100 URLs or more produce another map; regions which contain fewer than 100 URLs produce a ranked list of URL summaries.

ET-Map starts with a the top-level map showing forty odd broad entertainment 'subject regions' represented by regularly shaped tiles. Each tile is a visual summary of a group of Web pages with similar content. These tiles are shaded different colours to differentiate them, while labels identify the subject of the tile and the number in brackets telling you how many individual Web page links it contains. ET-Map can be browsed interactively, explored and queried, using the familiar point-and-click navigation style of the Web to find information of interest.

ET-Map was created using a sophisticated AI technique called Kohonen self-organizing map (SOM). See next project.

WEBSOM



Research Center