UNIT-5

Information Visualization

Contents

Information Visualization: What Is Infovis, Table Visualization: Printing the contents, Mapping values, Sampling issues.
Visualization of Relations: Tree Visualization,
Multivariate Data Visualization: Parallel Coordinate Plots, Dimensionality Reduction
Text Visualization: Content-Based Visualization, Visualizing Program Code

Infovis: Information Visualization

Datasets fields: Engineering, computational fluid mechanics and mathematics to medical and Earth sciences.

What Is Infovis

The field that studies the visual representation of database tables, text, and computer software data is known as information visualization (infovis)

How can we assist users in understanding all that abstract data?

Banking, telecom, and the information technology (IT) field to the logistics and administrative departments of large companies perform various activities in their respective industry generate a huge amount of data.A broad definition of infovis is "visualization applied to abstract quantities and relations in order to get insight in the data"

SciVis

The field that studies the visualization of these data types and targets the aforementioned application domains is known as scientific visualization (scivis).Scivis applications mainly focus on so-called **physical data**, which has an inherent spatial placement, such as the flow of water in a 3D container or a medical scan of a patient limb.In such cases, the user already has a mental image of what the flow container or the limb looks like.

InfoVis

SciVis has physical data whereas infovis has no inherent physical or spatial placement. Information has no shape or color and visualization is purely character. Infovis covers visual reasoning, visual data modelling, visual programming, visual information retrieval and browsing, visualization of program execution, visual languages, visual interface design, and spatial reasoning. Infovis : challenge of finding appropriate visual representations for such data. The three elements— representation, presentation, and interaction—form the fundamental ingredients of an infovis application. As a general rule, the design of infovis applications for any particular field should

- Follow the conventions accepted by that field;
- Integrate with other tools-of-the-trade of the field.

Table Visualization

The first type of infovis data is the table. For example, a database consists of a set of tables, each being a two-dimensional array of rows and columns. Each column typically describes a separate attribute, which is instantiated on each row.The table cells can contain all the attribute types. Databases are managed by means of interactive front-ends that allow users to perform query and editing operations on the tables. Also draw the contents of their tables in a two-dimensional grid of textual cell values. Oracle and Microsoft SQL Server, and open-source such as PostgreSQL and MySQL.

id	category	name	date	time	open	high	low	close
636	sF	SIF1	2004-11-29	13:00	0.800000	0.800000	0.800000	0.800000
635	sF	SIF1	2004-11-29	14:00	0.800000	0.800000	0.800000	0.800000
633	sF	SIF1	2004-11-29	16:00	0.795000	0.795000	0.795000	0.795000
630	sF	SIF1	2004-11-30	14:00	0.795000	0.795000	0.795000	0.795000
632	sF	SIF1	2004-11-30	12:00	0.800000	0.800000	0.795000	0.795000
631	sF	SIF1	2004-11-30	13:00	0.795000	0.795000	0.795000	0.795000
628	sF	SIF1	2004-11-30	16:00	0.795000	0.795000	0.795000	0.795000
629	sF	SIF1	2004-11-30	15:00	0.795000	0.795000	0.795000	0.795000
627	sF	SIF1	2005-00-02	12:00	0.795000	0.790000	0.785000	0.790000
626	sF	SIF1	2005-00-02	13:00	0.790000	0.795000	0.790000	0.795000
625	sF	SIF1	2005-00-02	14:00	0.795000	0.795000	0.795000	0.795000
624	sF	SIF1	2005-00-02	15:00	0.800000	0.800000	0.800000	0.800000
620	sF	SIF1	2005-00-03	15:00	0.795000	0.795000	0.795000	0.795000
623	sF	SIF1	2005-00-03	12:00	0.795000	0.795000	0.795000	0.795000
622	sF	SIF1	2005-00-03	13:00	0.795000	0.795000	0.795000	0.795000
621	sF	SIF1	2005-00-03	14:00	0.795000	0.795000	0.795000	0.795000
619	sF	SIF1	2005-00-03	16:00	0.795000	0.795000	0.795000	0.795000
618	sF	SIF1	2005-00-06	11:00	0.790000	0.790000	0.790000	0.790000
614	sF	SIF1	2005-00-06	15:00	0.795000	0.795000	0.795000	0.795000
617	sF	SIF1	2005-00-06	12:00	0.795000	0.795000	0.795000	0.795000
616	sF	SIF1	2005-00-06	13:00	0.795000	0.795000	0.795000	0.795000
615	sF	SIF1	2005-00-06	14:00	0.795000	0.795000	0.795000	0.795000
613	sF	SIF1	2005-00-06	16:00	0.795000	0.795000	0.795000	0.795000
609	sF	SIF1	2005-00-07	14:00	0.790000	0.795000	0.790000	0.795000
612	sF	SIF1	2005-00-07	11:00	0.795000	0.795000	0.795000	0.795000
611	sF	SIF1	2005-00-07	12:00	0.795000	0.795000	0.795000	0.795000
610	sF	SIF1	2005-00-07	13:00	0.790000	0.790000	0.790000	0.790000
608	sF	SIF1	2005-00-07	15:00	0.790000	0.790000	0.790000	0.790000
606	sF	SIF1	2005-00-08	13:00	0.795000	0.795000	0.795000	0.795000
607	sF	SIF1	2005-00-08	12:00	0.790000	0.790000	0.790000	0.790000
605	sF	SIF1	2005-00-08	14:00	0.795000	0.795000	0.795000	0.795000

Printing Contents

The simplest way to visualize a data table is to print its contents.

First, printing cannot show more than a few tens of rows and columns simultaneously. Second, every table row contains information recorded about the trading of a share during a certain time interval.

Stock brokers are interested in analysing this data in order to find trends, outliers, and correlations between the various recorded items, such as share prices, periods, and companies.

Improvements for effective Visualization

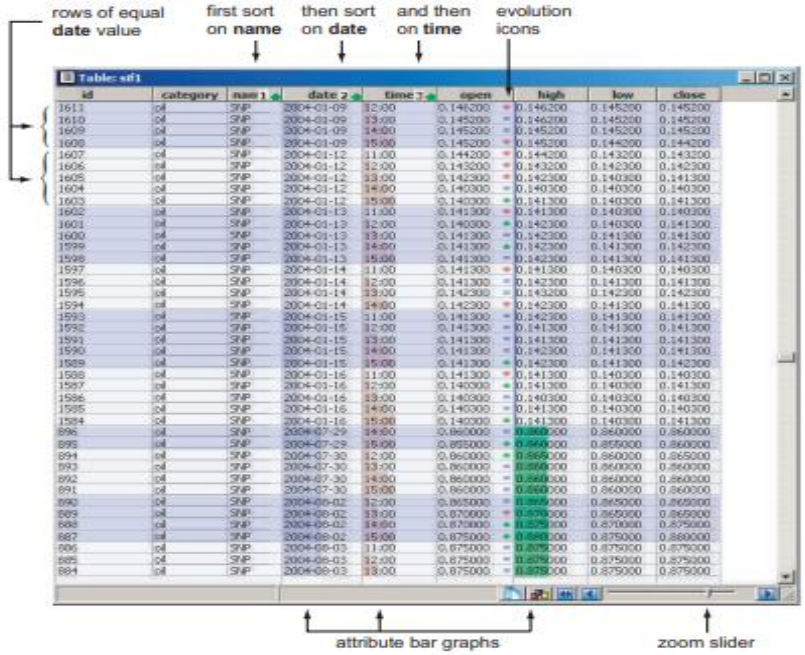
A first improvement is to sort the columns by their attribute value. This is useful to find out the range of the attribute, by examining the beginning and end of a sorted column.

Example: Show the evolution in time of the price of a given share.

- First, we sort the table rows by share name, which groups together the records (rows) corresponding to the same share.
- Next, we sort the same-name row groups by date.
- Finally sort the same-date row groups created by the second sort operation by time.

Three other enhancements besides the multiple sorting technique.

- The first enhancement helps us see how many intraday samples the dataset has. To do this, we draw the background of the table cells using alternating colors (light blue and white) that change between row groups belonging to different dates. We can now easily see that most trading days have four or five intraday samples.



- A second use of this visual cue is to facilitate focusing on the first and last intraday samples, which are important to assess a day's gains or losses, and also to compare the close of a day with the open of the next day.

Mapping Values

The second enhancement helps us follow the evolutions of the cell values without having to read them. Blend each cell's background with a colored bar graph showing the actual cell value. Short bars indicate small values, whereas a bar of the entire cell's length indicates the maximum value over the respective column. The third enhancement helps us follow the course variations during the monitored intervals.. A green upward-pointing arrow indicates a course increase, a red downward pointing arrow shows a course decrease, and a blue equal-sign icon shows an unchanged course.

Sampling issues

A different problem of the text-based table visualization is its limited scalability. Scrolling does not offer an overview of the entire data. Also, scrolling a table that has tens of thousands of rows or more is quite cumbersome. To solve this problem, we reduce the level of detail at which the table is shown, by zooming out the table visualization.

Visualisation of relations

Relation is an association between two or more items.Relational data is different from value data, as the information is located not in a single data value, but association of such values.

Applications

World Wide Web, cities connected by roads,on maps, suppliers and customers connected in a logistics network, and software components depending on each other in a software system architecture

Tree Visualization

Trees are a particular type of relational data.Formally, a tree T = (N,E) is defined as a set of nodes N = {ni} (also called vertices) and a set of edges.E = $\{e_i\}$, where every edge $e_i = (n_j , n_k)$ is a pair of nodes $n_j \in N$ and $n_k \in N$.

Property of a tree

There is a unique path, defined as a set of nodes connected by edges, between any two nodes in the tree.

A tree is a network of connected nodes where there are no loops. An edge $e_i = (n_j , n_k)$, the first node (n_j) is called the parent of n_k and the second one (n_k) is called the child of n_j .A tree can be seen as a hierarchical structure of parent and child nodes. A parent node may have any number of children, but a child node can have only one parent. This node is called the tree root, and represents the top level of the hierarchy encoded by the tree. A tree that have no children These nodes are called leaves, the bottom-most level of the hierarchy encoded by the tree. Finally, the depth of a tree is the length of the longest path (number of nodes) that connects a leaf to the root.

One such relation is containment, where parent nodes are seen as containers of child nodes.

Examples of containment tree hierarchies are

- Computer file systems,(files in folders),
- The structure of software source code (statements in functions
- In classes in files), and
- The logical map of a store (products in boxes in shelves in storage rooms).
- Another hierarchical relation is subordination, where parent nodes are seen as controllers of their children.
- Examples of subordination tree hierarchies are the structures of organizations (employees, managers, executives),
- the control structure of mechanical assemblies (driven parts connected to controllers),
- or the electrical network in a house (devices, power sockets, central electricity meter)

Multivariate Data Visualization:

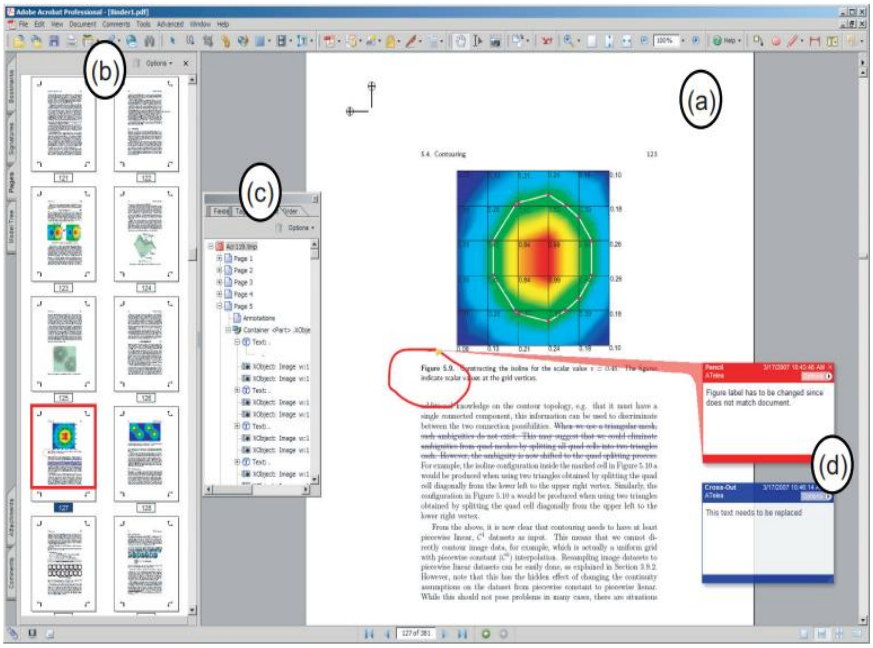
Consider a set of N data points D = {pi}, $1 \le i \le N$, where every point pi has a K-dimensional vector of attributes (a1i ,...,aKi) $\in$ AK, each attribute being defined over some domain A. Such a dataset is called multivariate, as it has several variables, or attributes, per data point

Parallel Coordinate Plots

The general problem of analyzing the distribution and correlation of positions of a set of K-dimensional points can be, conceptually, split into K problems of analyzing the distributions and correlations of K sets of one-dimensional values.We have a dataset containing around N = 400 data points. Each data point describes a car via K = 7 attributes. These data tuples can be seen as data points in a K = 7-dimensional space.Since we cannot directly render into seven dimensions, we must find ways to map this space onto two or three dimensions. One way to do this is to consider the K sets D, each containing N data values, as the K columns of a data table. A point p_j as a horizontal row, parallel coordinates. Map each point to a polyline that connects the points on the vertical axes whose ordinates (y values) equal the point attributes a_{ij} .

Text Visualization

Different types of information are contained in text document. Information can be structured into three categories: content, structure, and metadata. The content describes the information contained in the text itself. Structure characterizes how the text is organized hierarchically into several levels of abstraction, such as paragraphs, sections, chapters, or elements of a document collection. Metadata stores information about the document itself rather than information about the document content. Metadata includes cross references, keywords, and indexes, as well as information on the document author, publisher, and publication date.



Visualization methods

That target metadata should provide insight in the metadata itself. Ways to correlate metadata with the document content and structure. The other dimension of text visualization concerns the origin of the data. Example: Google search tool, analyze documents to generate information used for classification and indexing. Text analysis includes techniques that range from neural networks and statistical analysis to lexical, syntactic, and semantic analysis and natural-language processing.

Content based visualization

All views use simple and familiar two-dimensional layout and mapping techniques, such as the direct rendering of the document pages, at full or diminished size, and the tree-browser metaphor. The semantics of colors are also very intuitive. These represent either actual data in the document or user choices in the annotation process. Combined with simple navigation and interaction, these techniques can be quickly learned and used by a wide range of users

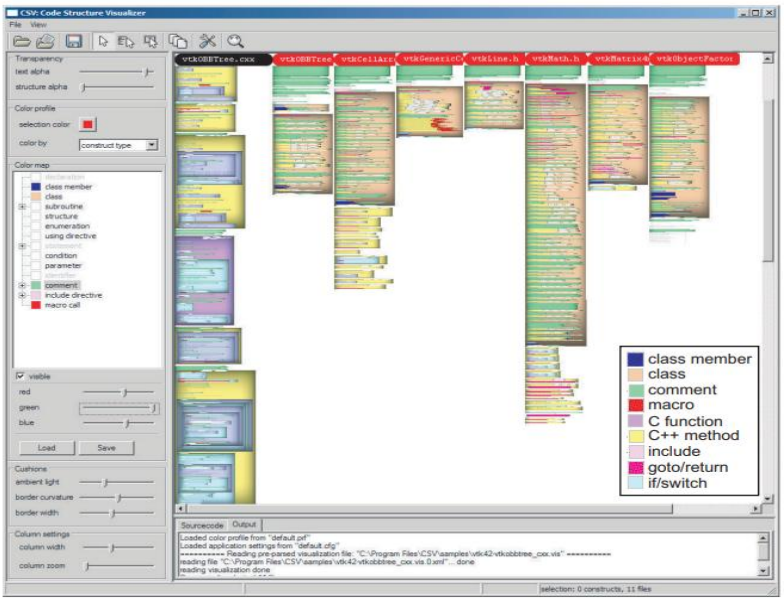
Visualizing Program Code

Modern systems have increasingly large sizes as measured in lines of code (LoC).
Source code has several particular properties, including the following:

• Exact: Source code is written in programming languages that have strictly defined grammars with non-ambiguous semantics.

• Large-scale: The source code of modern software systems has tens or thousands up to millions of lines of code.

- Relational: Source code contains many kinds of relations, such as the types of variables, members and parents of classes, dependencies of packages, clients of services, and interfaces of modules.
- Hierarchical: Source code contains many types of hierarchies, such as the package-file-class-method-statement hierarchy or hierarchies of data structures.



C source code visualization constructed with SeeSoft, one of the first tools to use the technique of mapping text lines to pixel lines. Several tens of files containing over five thousand lines of code in total. Color shows code age for each line: red shows recently modified lines, while blue shows lines unchanged for a long time.

- First, we get an overview of the relative sizes of all files in the project.
- Second, we quickly locate stable pieces of code that have not been changed for a long time, as well as recently changed code

Use of the same technique to gain a different type of insight into C++ source code.Layout of the original source code is used in a zoomed-out fashion. The user selects a number of syntactic structures of interest, such as functions, methods, classes, macros, includes, and conditional and jump statements, and assigns custom colors to them. For every such structure present in the source code, its outline is computed. The source code visualized in Figure is a part of the VTK class library. Each column depicts a separate file. The main advantages of showing source code in this way is the high scalability of the method and its intuitiveness given by using the original code layout.