Éligible à une allocation de type : UCA-EDSTIC-EUR-DS4H

Titre du sujet : Navigation and Visualization of Multidimensional Datasets over the Web: a visual approach for chained queries

Mention de thèse : Informatique

HDR Directeur de thèse inscrit à l'ED STIC : Marco Winckler

Co-encadrant de thèse éventuel :
Nom :
Prénom :
Email :
Téléphone :

Email de contact pour ce sujet : winckler@unice.fr

Laboratoire d'accueil : I3S

Description du sujet :
Information Visualization is an important aspect of data analysis that allows conveying information in a visual format highlighting patterns, trends, and correlations among data. They have been demonstrated as a suitable alternative for helping users to gain knowledge about the internal structure of abstract data and causal relationships in it because users can explore interactively the visual representation of datasets. Moreover, they are powerful tools specifically designed to support the exploration and analysis of large datasets, helping users to deal with complex decision-making tasks. Information visualization techniques might improve both users’ cognitive abilities and users’ performance with tasks by relieving the working memory and improving decision accuracy, even on elderly people. Many visualization techniques are intended to be used by specialized users but more and more often information visualization techniques are developed to larger audiences. Guided by appropriated tools, end-users can formulate visual queries to extract and visualize information from datasets publically available over the Web.
Many techniques for information visualizing exist. Graph visualization aims at showing the connectivity between data and are particularly relevant for the visualization of multidimensional datasets where each node in the graph content multiple attributes. The visualization of multidimensional data raises a dilemma between the benefits of showing as much as possible at once, to minimize the need for navigation and exploration, and the costs of showing too much at once, where the user is overwhelmed by visual clutter. Considering a large number of possible visualization techniques it is not possible to find a single combination of techniques that fully supports the exploration of these complex datasets. Thus, the solution seems to be providing the user with diverse techniques to be used in a combined way for the exploration of the data either a graph or a multivariate dataset.

In the context of the increasing amount of data source available on the Web, information visualization techniques might play a role in helping many users to explore data sources, understand data, and to support the construction of new knowledge. However, the inner connectivity of the Web implies that the exploration of complex and/or large datasets can only be achieved by triggering chained queries. In some cases, chained queries are simple ancillary search tasks, which are aimed at providing details on demand about the results that are currently in the display. However, in many cases, the exploration of datasets will require long sessions, for that user must be provided with a way of keeping track the exploration path, i.e., the tool must support provenance in both visualization and analysis.

This PhD thesis aims at investigating the use of information visualization techniques for supporting chained queries in datasets over the Web. The ultimate goal is to design, implement and assess an information visualization framework that would be able to support visual queries over the Web that can be chained in a sequence of visualization, thus representing the inner navigation/exploration of datasets by users. In a typical scenario, a user should be able to formulate a visual query over a Web dataset. The results will be shown in a visual format showing the correlation between data attributes. Users can then, select entries in the results of the initial query to create new queries whose results will be displayed in a new instance of the information visualization technique. The sequence of queries will create a graph that indicates the exploration of the information space. Each visual query should allow users to select a dataset over thus, which might allow users to navigate among multiple datasets using the same information visualization tool. This search pattern might have multiple applications in queries performed over the Web.

As we shall see, the construction of this framework implies many research questions of both scientific and pragmatic relevance. First, it will lead to the development of a structural model for supporting the construction of visual queries over the Web. Another interesting question is how to support chained queries over several types of datasets available over the Web, including semi-structured data and linked data. Linked data are of particular interest as a case study because it allows extracting semantically relevant attributes from chained queries. In this context, information visualization techniques can also help to develop new strategies for extracting knowledge from linked datasets. Another interesting research question refers to the
use of visual chained queries by end users. This question might understand how users perceive and cognitively process information and how they make use of tools to find the information they need and weather (or not) they can make use of these tools to create new knowledge.

This PhD thesis is part of a project that aims at foster the collaboration with the INRIA team WIMMICS (for research related to the exploration of linked datasets) and the Universidad National de La Plata (Argentina) (for the aspects related to Web Engineering).

**URL:** [https://drive.google.com/file/d/1M6afBI0T1tMvNnTLLNtx4hncnYZ2QDUDa/view?usp=sharing](https://drive.google.com/file/d/1M6afBI0T1tMvNnTLLNtx4hncnYZ2QDUDa/view?usp=sharing)

**English version:**