

Import automatique et interactif de données dans les systèmes de visualisations

Frédéric Gilbert*, David Auber*

*LaBRI, Université de Bordeaux, Talence, France
Gravité Inria Bordeaux Sud-Ouest
frederic.gilbert@labri.fr,
david.auber@labri.fr

Résumé. La première étape du processus de visualisation d'information consiste à transformer les données d'un format brut vers une structure de données utilisable par les différents composants de visualisation. Dans les applications réelles, cette première étape représente une barrière empêchant l'accès des utilisateurs novices à une riche variété de techniques de visualisation. Par exemple, il peut être techniquement impossible pour un utilisateur lambda de transformer des données arborescentes en un modèle de graphe pouvant utiliser une représentation à base de TreeMap. Une autre barrière est aussi la multitude de transformations possible des données brutes. Il faut pouvoir explorer cet ensemble de combinaisons. Basé sur nos retours d'expériences avec des utilisateurs finaux, dans cet article, nous considérons que le format brut est sous forme tabulaire. Ce format est le plus couramment utilisé et est facilement accessible par nos utilisateurs. Nous proposons une méthode novatrice permettant de générer automatiquement des graphes valués à partir de n'importe quelle table. En analysant le contenu de chaque dimension nous identifions les interconnexions entre celles-ci. Puis nous caractérisons les entités, les attributs et les relations possibles au sein des tables. Finalement, nous intégrons l'utilisateur dans le processus de transformation en lui proposant un ensemble de transformations valides.

1 Introduction

De nos jours, la collecte de données est de plus en plus aisée. Il existe un grand nombre d'outils et méthodes pour collecter et stocker facilement des données. Le plus souvent ces données sont stockées sous forme de base de données contenant des informations basiques comme des noms, des nombres, des dates, des heures... Elles sont stockées dans le but de garder des informations comme qui/quoi, quand, où, etc... Mais ces facilités de collecte et stockage ont pour conséquences de voir augmenter la quantité de grandes bases de données. Par exemple, les chercheurs sur le génome humain collectent des données dans le but de comprendre comment les gènes interagissent et les compagnies commerciales en collectent afin de mettre en évidence des comportements typiques ou atypiques de clients.

Pour répondre à ce type de questions, il existe des systèmes de visualisation comme DE-Vise Livny et al. (1997), Polaris Stolte et al. (2008) ou Tableau Mackinlay et al. (2007). Ces