Visually Supporting Data Mining of Large Existing Databases

Daniel A. Keim, Hans-Peter Kriegel

Institute for Computer Science, University of Munich
Leopoldstr. 11B, D-80802 Munich

Abstract

In this paper, we present ideas how visualization technology can be used to improve the difficult process of querying very large databases. With our VisDB system, we try to provide visual support not only for the query specification process, but also for evaluating query results and, thereafter, refining the query accordingly. The main idea of our system is to represent as many data items as possible by the pixels of the display device. By arranging and coloring the pixels according to the relevance for the query, the user gets a visual impression of the resulting data set and of its relevance for the query. Using an interactive query interface, the user may change the query dynamically and receives immediate feedback by the visual representation of the resulting data set. By using multiple windows for different parts of the query, the user gets visual feedback for each part of the query and, therefore, may easier understand the overall result. To support complex queries, we introduce the notion of 'approximate joins' which allow the user to find data items that only approximately fulfill join conditions. We also present ideas how our technique may be extended to support the interoperation of heterogeneous databases. Finally, we discuss the performance problems that are caused by interfacing to existing database systems and present ideas to solve these problems by using data structures supporting a multidimensional search of the database.

Keywords: Data Mining, Visual Query Systems, Visual Relevance Feedback, Interfaces to Database Systems, Visualizing Large Data Sets, Visualizing Multidimensional and Multivariate Data, Interface and Visualization Technology