



The prototype of the DARE system

Tiziana Catarci, Giuseppe Santucci
Dipartimento di Informatica e Sistemistica
Università di Roma "La Sapienza"
Via Salaria, 113 - Roma, Italy
[catarci/santucci]@dis.uniroma1.it

When building a database, it is mandatory to design a friendly interface, which allows the final user to easily access the data of interest. Very often, such an interface exploits the power of visualization and direct manipulation mechanisms. However, it is not sufficient to associate "any" visual representation to a database, but the visual representation should be carefully chosen to effectively convey all and only the database information content.

We are currently working on a general theory (see ¹) for establishing the adequacy of a visual representation, once specified the database characteristics, and we are developing a system, called *DARE: Drawing Adequate REpresentations*, which implements such a theory. The DARE system is based on a knowledge base containing different kinds of rules, namely: *Visual rules*, that characterize the different kinds of visual symbols; *Data rules*, that specify the characteristics of the data model, the database schema, and the database instances; *Mapping rules*, that specify the link between data and visual elements; *Perceptual rules*, that tell us how the user perceives a visual symbol (e.g., a line), relationships between symbols (e.g., the mutual placements of two figures), and which is the perceptual effect of relevant visual attributes such as color, texture, etc.

The prototype

The first DARE prototype has been implemented to allow the user analyzing large amounts of data representing query results. Several attempts have been made to deal with such a problem, most of them focusing on the idea of visually refining the initial query, allowing the user to restrict the final result by changing the value of the attributes through suitable widgets. The problem is that such approaches have been manually tailored for specific applications and it is very difficult to generalize them. Furthermore, the user is typically not provided with a feedback about the selectivity of the attributes involved in the query. This is an important

¹T. Catarci, M.F. Costabile, G. Santucci. Foundations of the DARE System for Drawing Adequate Representations. In *Proc. of DANTE'99*, Japan, 1999.

Permission to make digital or hard copies of part or all of this work or personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee.

ACM SIGMOD 2001 May 21-24, Santa Barbara, California USA
Copyright 2001 ACM 1-58113-332-4/01/05...\$5.00

issue, since it could happen that a little variation on the selection conditions results in a great reduction of the result cardinality.

In order to build an effective system we made some assumptions, namely: to use the relational model, to consider queries that consist of simple selections, to limit the number of attributes involved in the query (e.g., 5). It is worth noting that these restrictions are not so severe: it is a well known fact that select-project queries represent the majority of the query asked by users and it is very seldom that the number of selection attributes is greater than three.

The problem can be defined more precisely as follows: having a relational table representing the result of a selection involving few attributes through range conditions, which are the most suitable visual symbols and visual variables able to effectively convey to the user the distribution of the tuples across the selection attribute domains?

Concerning the visual symbols, it is quite evident that the point is the best candidate to represent thousands of tuples on the same screen. The visual attributes available for the point are the spacial coordinates, the color hue, and the color lightness, resulting in five or four different attributes depending on whether we consider or not 3D representations. On the other hand, if the number of tuples to be represented is less than a threshold value (wrt screen resolution and dimension) a tuple can be represented by a 2D figure characterized by further visual attributes.

In the prototype a suitable set of rules, covering a subset of the overall DARE framework, has been implemented, namely: visual rules, characterizing points and simple 2D/3D figures; data rules, associated with the concepts of relation, tuple, attribute, and domain; mapping rules, specifying the association among the relational attributes and the visual attributes; and perceptual rules, concerning the best representation of relational attributes. Using such rules and based on the query result cardinality, the system is able to automatically choose the suitable visual elements to adopt to represent the tuples and the best visual representation for the attributes involved in the query. The result representation can be further manipulated by the user. It is possible to pick up with the mouse a single point and use sliders to restrict the result. Moreover, several "visual data mining" primitives are available and it is possible to project data on the three planes or to cluster them in order to visually capture data distribution and data relationships. The prototype, written in Java 1.3, uses the Magician 2.1.0 libraries and allows for querying (through JDBC) any relational table or to search the Web (e.g., through Google).