OLAP preferences: A research agenda

Stefano Rizzi (Univ. of Bologna - Italy)

Panel: Research Challenges for DW and OLAP Seen From Industry and Academia

Motivation

- Personalizing e-services by allowing users to express preferences is becoming more and more common
  - ✓ When querying, expressing preferences is seen as a natural way to avoid empty results on the one hand, information flooding on the other
  - ✓ Besides, preferences allow for ranking query results so that the user may first see the data that best match his tastes
- No attempt to develop a theory of preferences on OLAP cubes has been made yet!

Example

- Consider a classical marketing scenario, where users covering different roles ask for accessing data in different contexts
  - ✓ The sales manager is mainly interested in monthly revenues, but he may also wish to see more detailed data if the weekly revenue exceeds 10 k Euro
  - ✓ A sales agent prefers to see highest commissions data for his own customers
  - ✓ A branch manager is more interested in products of a given category selling abnormally low quantities
- Formulating such queries by expressing hard constraints on the cube data, would probably lead either to information flooding or to empty results
- Formulating the same queries by expressing preferences, meant as soft constraints on data, could improve the user satisfaction by returning the information that achieves the best compromise between his wishes

Basics on preferences

Quantitative preferences indirectly expressed by means of a scoring function that associates a numerical score to each tuple returned by a query

Qualitative preferences

directly expressed as binary relations on the space of tuples

Kiessling approach

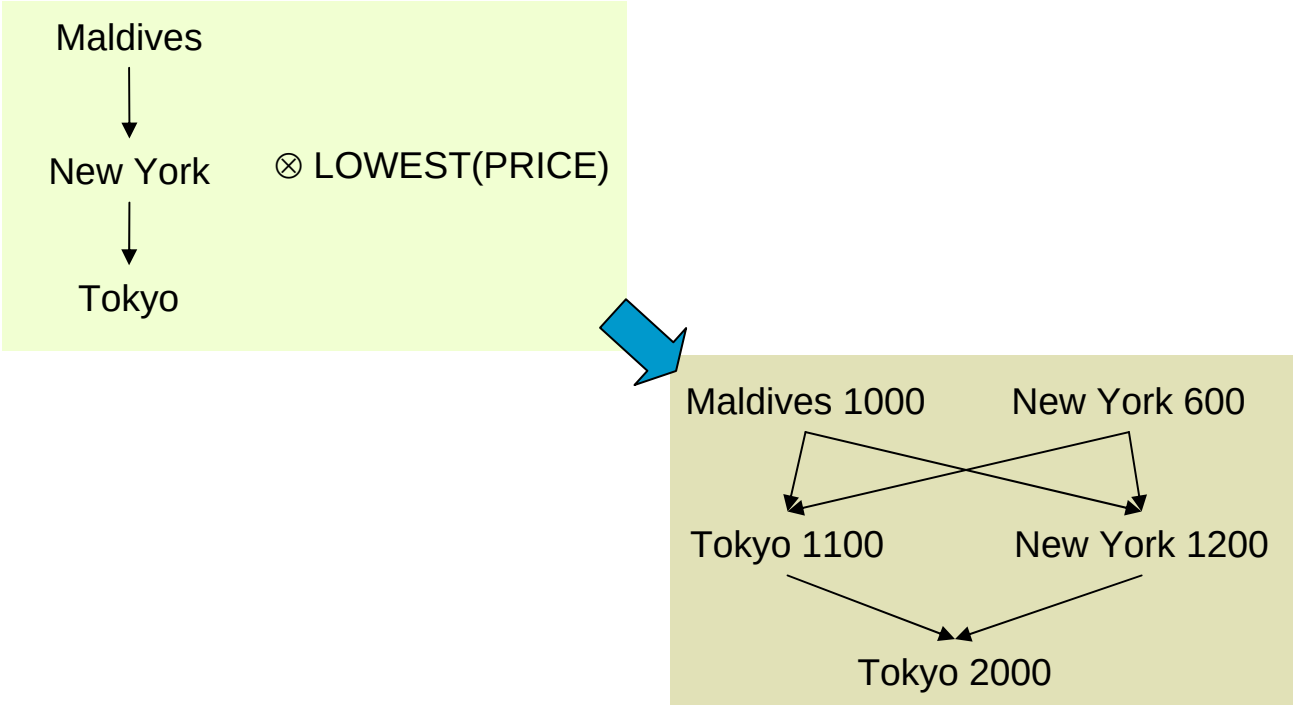
complex preferences are inductively constructed by applying composition operators to a set of predefined *base preference constructors*, thus obtaining a preference algebra

Chomicki approach

preferences are ``freely'' defined by formulating first-order *preference formulas*

Basics on preferences

- A preference is a strict partial order (an irreflexive and transitive binary relation) over the set of all possible combinations of attribute values



A research agenda

- An ad hoc approach must be devised for dealing with preferences on OLAP cubes
- Some research issues to be faced:
  - ✓ Preference model
    - The aggregation level has a strong impact on the size of the result returned to the user
    - In the OLAP domain, users must be enabled to express their preferences on the query aggregation level too, for instance by stating that monthly data are preferred to yearly and daily data
  - ✓ Context-awareness
    - Within a context-aware preference system, in an OLAP ubiquitous setting, the context could be determined by a set of coordinates such as the user role, the type of device he is operating, the type of information he wants to analyze, and his spatio-temporal location

A research agenda

- ✓ Query optimization
  - In presence of complex preferences, it becomes essential to rely on a set of equivalence rules to be used for rewriting expressions in a better performing way
  - Introducing operators for expressing preferences on the aggregation level will open new optimization possibilities
- ✓ Query processing
  - It is necessary to develop original processing techniques capable of efficiently coping with preferences on aggregation levels by relying on ad hoc algorithms and, possibly, new types of indexes
- ✓ User interface
  - How to enable users to express preferences through an OLAP front-end?
  - If preferences are to be also expressed on the aggregation level, data with different granularities may be returned together as the result of a query, which makes the traditional tabular and diagrammatic forms for viewing results fall short

Example revisited

- A sales agent prefers to see highest, preferably monthly, commissions data for his own customers

POS(Agent, ‘Smith’) ⊗ HIGHEST(Revenue) ⊗ NEAR(Time, Month)

- A branch manager is more interested in products of a given category and/or selling abnormally low quantities

POS(Category, ‘Video’) ▷ LOWEST(Revenue)