Determining Preferences from Semantic Metadata in OLAP Reporting Tool

Darja Solodovnikova, Natalija Kozmina

Faculty of Computing, University of Latvia, Riga LV-1586, Latvia
{darja.solodovnikova, natalija.kozmina}@lu.lv

Abstract. In this paper we present a reporting tool, which is a part of the data warehouse framework developed and put into operation at the University of Latvia, from a viewpoint of different metadata layers. All operation of the data warehouse framework and the reporting tool is based on metadata that consists of five interconnected parts, which are exposed in this paper: logical, physical, semantic, reporting, and OLAP preferences metadata. User OLAP preferences are essential for generating recommendations on potentially interesting reports, therefore we considered the process of semantic metadata usage at the stage of formulating user preferences.

Keywords: data warehouse, user preferences, business metadata, reports

1 Introduction and Related Work

Sometimes, during sessions of work with a reporting tool, a user has no notion about what kind of data he/she is able to find there. Moreover, a user might be unaware of a potentially useful report, because, for instance, it has been created recently and the user hasn't examined it yet. In one of our works [1] we focused on acquiring user preferences implicitly either by analyzing his/her previous activities or by learning the structure of the browsed report in order to suggest him/her other reports that might be helpful, meanwhile saving user's time and effort. In this paper we concentrate on preferences explicitly formulated by users of the OLAP reporting tool.

Apart from employing the reporting tool as a means of creating, modifying and executing reports on data warehouse schema, we also consider this reporting tool as an experimental environment for introducing OLAP personalization. Users of the reporting tool may have different skill levels (e.g., expert, novice), that's why reports' recommendations based on user preferences are more valuable for novice users than for experts. The reporting tool is a part of the data warehouse framework [2] developed at the University of Latvia.

The ideas of introducing personalization into data warehouses came from the field of databases [3] and still remain a subject of interest. Data warehouse can be personalized at schema level, applying rules for the data warehouse personalization, thus, giving a user an opportunity to work with a personalized OLAP schema, which matches his/her needs [4]. Users may express their preferences on OLAP queries [5];

in such case, the problem of performing time-consuming OLAP operations to find the necessary data is significantly improved. The other method of personalizing OLAP systems is to provide query recommendations to data warehouse users via investigating former sessions of the same user [6], or via collecting user preferences into a profile and processing it, while generating query recommendations [7]. Another aspect of OLAP personalization is the visual representation of data [8]: multiple layouts and visualization techniques may be interactively used for various analysis tasks. The summary of the research made in the field of personalization in OLAP is found in our previous work [9].

The rest of the paper is organized as follows: section 2 introduces interrelated metadata layers of the reporting tool, section 3 describes the process of user preference formulation in business language and its further transformation, and section 4 concludes the paper.

2 OLAP Reporting Tool

All operation of the data warehouse framework and the OLAP reporting tool as a part of it is based on metadata that consists of five interconnected parts (fig. 1). Logical metadata is used to describe data warehouse schemata, physical metadata describes storage of a data warehouse in relational database, semantic metadata describes data stored in a data warehouse and data warehouse elements in a way that is understandable to users, reporting metadata stores definitions of reports on data warehouse schemata, and OLAP preferences metadata stores definitions of user preferences on reports' structure and data.

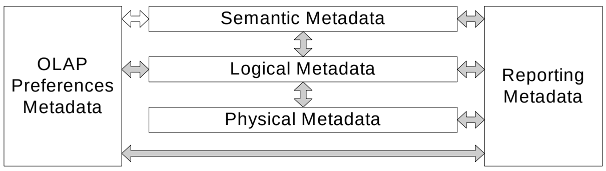

Fig. 1. Metadata connections

Particular classes of parts of metadata are connected by associations. Semantic metadata describes report items from the reporting metadata and data warehouse schema elements from the logical metadata. Data warehouse schema elements from the logical metadata correspond to tables and table columns described in the physical metadata. Items of reports defined in the reporting metadata are obtained from table columns described in the physical metadata and correspond to data warehouse schema elements from the logical metadata. OLAP preferences metadata defines user preferences for data warehouse schema elements described in the logical metadata and for reports described in the reporting metadata. OLAP preferences are formally defined by concepts of semantic metadata. To be more precise, components of user preferences on reports' structure are OLAP schema elements from the logical metadata that correspond to concepts from the semantic metadata, and components of user preferences on reports' data are items of reports from the reporting metadata that are defined by concepts as well. Thereby, there is a latent connection between semantic metadata and OLAP preferences metadata.

Common Warehouse Metamodel (CWM) [10] was used as a basis for the semantic, logical and physical metadata. CWM is a metadata standard produced by Object Management Group to simplify metadata interchange between data warehousing applications. CWM consists of packages, which describe different aspects of a data warehouse.

2.1 Logical Metadata

Metadata at the logical level describes the multidimensional data warehouse schema (fig. 2.). The logical level metadata is based on the OLAP package of Common Warehouse Metamodel (CWM) [10] and contains the main objects from this package such as dimensions with attributes and hierarchies, fact tables (cubes in CWM) with measures. Fact tables and dimensions are connected by FactTableDimension associations. Only dimensions and fact tables connected by FactTableDimension associations can be included together in one report. OLAP package of CWM was extended by the class AcceptableAggregation, which stores information about aggregate functions (SUM, AVG, COUNT, MIN, MAX) acceptable for each measure and dimension. This metadata is essential for correct queries. The detailed description of all metadata levels of a data warehouse, including the description of the logical level, is found in the paper [11].

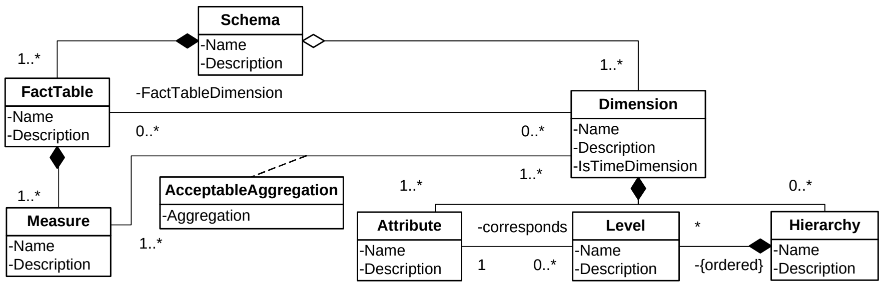

Fig. 2. Logical level metadata [11]

2.2 Reporting Metadata

Reporting metadata describes the structure of reports on data warehouse elements (fig. 3). Basically, reports are worksheets that contain data items defined by calculations, which specify computation formulas from parameters and table columns that usually correspond to schema elements (measures and attributes). Reports also consist of user-defined conditions and joins between tables.

Reports in the tool are defined by developers or experienced users themselves by choosing desired elements of a data warehouse schema and defining conditions, parameters, etc. To define one report, users are allowed to select measures and attributes belonging to one schema. According to the report definition, reporting metadata is created for each report. When a user runs a report in the OLAP reporting tool, an SQL query is built based on the report definition in reporting metadata [12], and its result is displayed to a user.

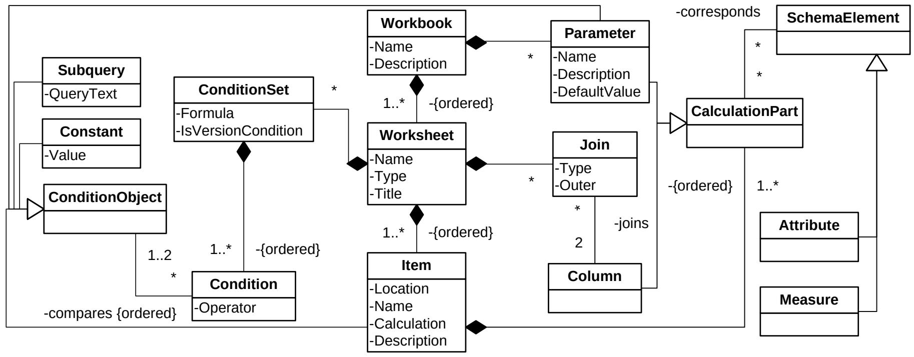

Fig. 3. Reporting metadata [12]

2.3 Semantic Metadata

It is essential for data warehouse users to understand the semantics of data that appears in reports from business perspective.

There are multiple reasons why it is necessary to describe each element of the data warehouse model in business language. For instance, while working with the reporting tool, users also must be able to analyze this data using all necessary features, including OLAP operations drill-down and roll-up and using hierarchies. Besides, it is desirable that users can modify or construct reports themselves from elements, which are familiar to them, so that reports' creation becomes transparent. Moreover, users should be able to state their OLAP preferences, operating with business language terms, so that it would be possible to provide users of different skill levels (e.g., expert, novice) with recommendations on potentially interesting reports. The description on business level could also be employed by users to express their requirements for information and also changes in requirements, thus, making the understanding between users and developers of data warehouse clearer.

Data warehouse elements' description in business language is stored in the semantic metadata.

In CWM there is the package Business Nomenclature, which can be used to represent business metadata. This package was taken as a basis for semantic metadata depicted in fig. 4. The main classes that are used for description of data warehouse elements are Terms and Concepts, which are united in Glossaries and Taxonomies respectively. A concept is the semantic meaning or notion of some data warehouse element or data stored in some element, but a term is particular word or phrase employed by users to refer to a concept. There may be preferred terms and synonym terms to identify a concept. Also, terms may be related to each other. In semantic metadata Concepts define elements of a data warehouse schema (class

SchemaElement from the logical metadata) and items used in reports (class Item from the reporting metadata).

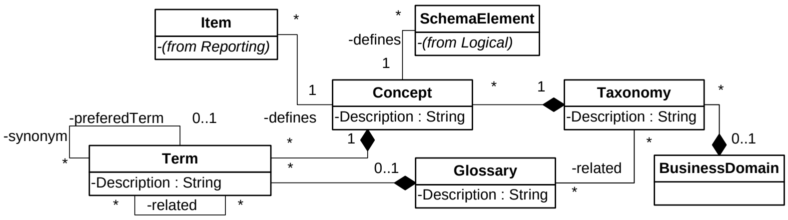

Fig. 4. Semantic metadata

2.4 OLAP Preferences Metadata

A metamodel, previous version of which is published in [13], describes OLAP schema preferences and is depicted in fig. 5. There are two kinds of user preferences: (i) Schema-Specific preferences on OLAP schema, its elements and acceptable aggregate functions, and (ii) Report-Specific preferences on data in reports.

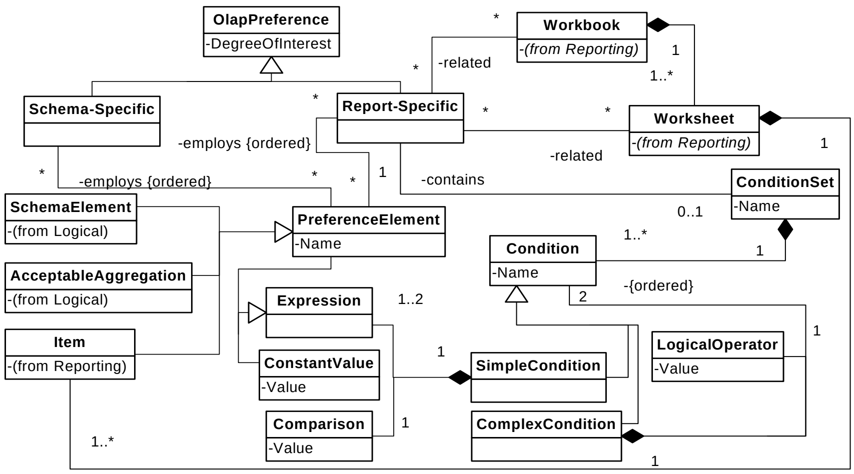

Fig. 5. OLAP preferences metadata

A user may set the degree of interest (DegreeOfInterest, *doi* [3]) for each OLAP preference. For instance, a user operates with values of the DegreeOfInterest attribute that may be the following: very low, low, medium, high, very high. Each degree of interest may have a defined real number equivalent from the interval [0; 1] that is assigned automatically. For example, medium degree of interest corresponds to the numeric value 0.5, low degree of interest – to 0.2, etc.

A schema-specific preference does not refer to a specific set of reports (i.e. workbook) or a single report (i.e. worksheet). However, it refers to OLAP schema as a whole.

A PreferenceElement class describes the type of the element in user preference, which may be an OLAP schema, an OLAP schema element (e.g. dimension, fact table, attribute, etc.), an aggregate function, or a report's item.

Unlike schema-specific preferences, the scope of report-specific preferences is either a worksheet or a workbook. In report-specific preferences one or more preference elements (Items) may be included, and vice versa, a single preference element (Item) may be used in multiple user preferences of that type.

Report-specific preferences include restrictions on report data. Each report-specific preference may contain a set of conditions (ConditionSet). A Condition class is divided into two subclasses: a SimpleCondition and a ComplexCondition. A complex condition consists of two or more simple conditions, joined with a logical operator (AND, OR). A simple condition consists of two expressions (Expression) and a comparison operator (Comparison). It is allowed to apply the following comparison operators: =, <>, >=, <=, >, <, IN/NOT IN, IS NULL/IS NOT NULL, LIKE/NOT LIKE, EXISTS/NOT EXISTS. Typically, one expression is a preference element and the other is a constant value (ConstantValue), which is either a string of symbols or a numeric value. There may be also just one expression, i.e. preference element, in case when the value of the comparison operator is EXISTS/NOT EXISTS or NULL/IS NOT NULL.

3 Determining Preferences from Semantic Description

We consider semantic metadata as means of formulating user preferences for data warehouse reports, applying pre-defined description of data warehouse elements.

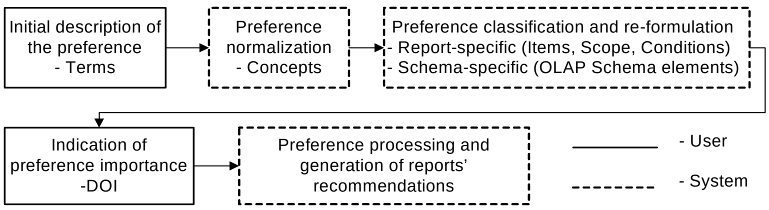

Fig. 6. Processing user preferences described with semantic metadata

The process of preference creation and transformation is briefly depicted in fig. 6, and is the following:

1) A user describes his/her preference, choosing one of the synonym terms from the glossary that seem to be the most suitable and understandable for him/her (fig. 4.). *Example*: terms "study program", "academic specialization", "branch", "field of study" are considered as synonyms, among which a user is free to select the most appropriate one.
2) A set of terms corresponds to exactly one concept. Thus, we normalize user preferences, transforming terms into concepts. *Example*: terms mentioned in 1) are all related to one concept, which is "study program".
3) The type of the user preference is being detected. Bearing in mind that each concept defines either report items or OLAP schema elements, user preferences are

later re-formulated, employing either items or OLAP schema elements instead of concepts (fig. 4.). If one concept corresponds to several OLAP schema elements or report items, then the number of preferences increases respectively.

a) For a preference to be classified as report-specific, a scope (worksheet or workbook) of the preference should be indicated, and, if necessary, a set of conditions should be created, employing elements from the metamodel in fig. 5. In such case, a concept is linked to a report's item. *Example*: a condition is a restriction on data, for instance, "study program name = "Information Systems" ".

b) A schema-specific preference is more general than a report-specific one; there are no scopes or conditions in it. In such case, a concept is linked to OLAP schema element.

4) In compliance with metamodel in fig. 5, a degree of interest should be assigned by user to each OLAP preference. *Example*: a medium degree of interest is equal to 0.5.
5) When all OLAP preferences are formed, schema- and report-specific preferences are processed separately in order to provide user with recommendations on reports.

a) The method for generating recommendations based on report-specific preferences has not been considered by authors yet and is a subject of future research.

b) However, there is a hot-start method for providing recommendations on reports based on implicitly discovered schema-specific preferences described in [1]. This method can be adapted and applied for explicitly discovered schema-specific preferences.

4 Conclusions and Future Work

In this paper we paid attention to a reporting tool, developed and currently being used at the University of Latvia. We exposed five different layers of metadata that intersect each other: logical metadata that describes data warehouse schemata, physical metadata that describes storage of a data warehouse in relational database, semantic metadata that describes data stored in a data warehouse and data warehouse elements in a way that is understandable to users, reporting metadata that stores definitions of reports on data warehouse schemata, and OLAP preferences metadata that stores definitions of user preferences on reports' structure and data.

We considered a possibility for a user to create OLAP preferences, using description in business language, operating with synonym terms and choosing the most appropriate among them. We briefly set forth a concept of the algorithm of OLAP preference creation, transformation and processing.

There are several directions of our future work; we would like to extend and supplement our algorithm of OLAP preference creation, transformation and processing, thus, leading it to the level, which is closer to implementation.

Also, we would like to review the existing approach for generation of reports' recommendations [1]. This approach is based on implicitly discovered schema-specific user preferences; however, it is worthwhile to adapt it to explicitly set user preferences. Along with that a method for handling report-specific user preferences should be developed. The evaluation of processing both types of explicitly set user preferences (schema- and report-specific) will follow.

Acknowledgments. This work has been supported by ESF project No. 2009/0216/1DP/1.1.1.2.0/09/APIA/VIAA/044.

References

1. Solodovnikova, D., Kozmina, N.: On Implicitly Discovered OLAP Schema-Specific Preferences in Reporting Tool. In: Proceedings of BIR2011 International Conference. To be published. (2011)
2. Solodovnikova, D.: Data Warehouse Evolution Framework. In: Proceedings of the Spring Young Researcher's Colloquium On Database and Information Systems (SYRCoDIS'07), Moscow, Russia (2007), http://ceur-ws.org/Vol-256/submission_4.pdf
3. Koutrika, G., Ioannidis, Y. E.: Personalization of queries in database systems. In Proceedings of 20th International Conference on Data Engineering (ICDE'04), Boston, MA, USA, pp. 597-608. (2004)
4. Garrigós, I., Pardillo, J., Mazón, J.-N., Trujillo, J.: A Conceptual Modeling Approach for OLAP Personalization. Conceptual Modeling - ER 2009, LNCS, vol. 5829, Springer, Heidelberg, pp. 401-414. (2009)
5. Biondi, P., Golfarelli, M., Rizzi, S.: Preference-Based Datacube Analysis with MyOLAP. In Proceedings of the 27th International Conference on Data Engineering (ICDE 2011), Hannover, Germany, pp. 1328-1331. (2011)
6. Giacometti, A., Marcel, P., Negre, E., Soulet, A.: Query Recommendations for OLAP Discovery Driven Analysis. In Proceedings of 12th ACM International Workshop on Data Warehousing and OLAP (DOLAP'09), Hong Kong, pp. 81-88. (2009)
7. Jerbi, H., Ravat, F., Teste, O., Zurfluh, G.: Preference-Based Recommendations for OLAP Analysis. In Proceedings of the 11th International Conference on Data Warehousing and Knowledge Discovery (DaWaK'09), Linz, Austria, pp. 467-478. (2009)
8. Mansmann, S., Scholl, M. H.: Visual OLAP: A New Paradigm for Exploring Multidimensonal Aggregates. In Proceedings of IADIS International Conference on Computer Graphics and Visualization (MCCSIS'08), Amsterdam, The Netherlands, pp. 59-66. (2008)
9. Kozmina, N., Niedrite, L.: Research Directions of OLAP Personalization. In: Proceedings of the 19th International Conference on Information Systems Development (ISD'10), Prague, Czech Republic, to be published. (2010)
10. Object Management Group: Common Warehouse Metamodel Specification, v1.1, http://www.omg.org/cgi-bin/doc?formal/03-03-02
11. Solodovnikova, D.: Metadata to Support Data Warehouse Evolution. In: Proceedings of the 17th International Conference on Information Systems Development (ISD'08), Paphos, Cyprus. pp. 627-635. (2008)
12. Solodovnikova, D: Building Queries on Multiple Versions of Data Warehouse. In: Haav, H.-M., Kalja, A. (eds), Databases and Information Systems V - Selected Papers from the 8th International Baltic Conference DBIS 2008, IOS Press, pp. 75–86. (2008)
13. Kozmina, N., Niedrite, L.: OLAP Personalization with User-Describing Profiles. In: Forbrig, P., Günther, H. (eds.) BIR 2010. LNBIP, vol. 64, Springer, Heidelberg, pp. 188-202. (2010)