Multidimensional Schemas Quality Assessment

Nicolas PRAT[1] and Samira SI-SAID CHERFI[2]

[1] ESSEC; Avenue Bernard Hirsch; BP 105; 95021 Cergy Cedex; FRANCE
prat@essec.fr
[2] CEDRIC-CNAM; 292, rue Saint-Martin ; 75141 Paris Cedex 03 ; FRANCE
sisaid@cnam.fr

Abstract: A data warehouse is a database focused on decision making. It is built separately from the transactional (OLTP) databases of the enterprise, although it is partly fed from transactional data. Data warehouses are typically accessed by decision makers using OLAP tools, based on a specific, multidimensional representation of data. Considering the strategic importance of data warehouses, the quality of these systems is crucial. Moreover, since OLAP tool users directly access multidimensional schemas, multidimensional schemas quality is a key aspect of data warehouse quality. This paper focuses on the quality of multidimensional schemas. We present the underlying multidimensional model and the multidimensional schemas quality evaluation framework, which considers three views corresponding to three types of users: the specification view (data warehouse designer), the usage view (decision maker) and the implementation view (data warehouse developer). Concentrating on the specification view, we present and exemplify criteria and metrics pertaining to this view.

1. Introduction

A data warehouse is a database aimed at decision making, built by integrating data from external sources and from internal OLTP (On-Line Transactional Processing) systems. OLAP (On-Line Analytical Processing) tools represent data in a multidimensional fashion, enabling business users to formulate queries and perform analyses.

Quality issues raised by data warehouses are crucial. Previous work on data warehouse quality has often focused on the key issue of data quality, in particular the quality of source transactional data. However, the quality of the data model, i.e. the evaluation of multidimensional schemas, is also a crucial issue, all the more so as in OLAP systems, users access data directly.

In this paper, we focus on multidimensional schemas quality evaluation. For the paper needs, we assume that a multidimensional schema is defined based on (1) the user requirements in terms of analysis (queries, lists of attributes, schemas modeled with the ER notation [1] etc.) and/or (2) the schema of operational data sources (represented with the ER or EER notation). Once the multidimensional schema has been defined, it may be implemented in an OLAP tool.

Since many multidimensional models have been defined in the literature [2,3] and no standard has emerged yet, our approach uses a unified multidimensional model based on previous work [4,5]. The model unifies and integrates the key concepts of the multidimensional models found in the literature.

Our approach for multidimensional schemas quality evaluation adapts the framework defined in the OLTP context for conceptual schemas quality evaluation [6,7]. We consider three viewpoints: the first viewpoint, named *specification*, is concerned with the data warehouse designer. The second viewpoint, called *usage*, considers the point of view of the data warehouse user i.e. the decision maker. Finally, the *implementation* viewpoint deals with physical issues and concerns the data warehouse/data mart developer. For each viewpoint, a set of criteria is defined, with associated metrics facilitating (semi)-automatic multidimensional schemas quality evaluation.

The paper is organized as follows. Section 2 describes related research. Section 3 presents the multidimensional model used in our approach. Section 4 presents the general framework for multidimensional schemas quality assessment and focuses more precisely on the specification view, presenting and illustrating a set of criteria and metrics pertaining to this view. Section 5 concludes and describes further research.

2. Related Research

Several approaches, which deal with the evaluation of software products, exist. They can be summarized according to how they consider different phases in software life-cycle such as software design, software development and maintenance, and data quality. A synthetic presentation of the literature related to the quality assessment can be found in [7].

Regarding quality evaluation in *data warehouse environments*, previous work can be classified into three categories :

- Due to the importance of operational data sources quality, and more generally *data quality in data warehouses*, many papers are dedicated to this issue. [8] proposes a risk-based approach to data quality assurance in data warehouses. [9] presents ideas and describes a model to support data quality enhancement in data warehouses.
- The second category includes research dedicated to *multidimensional schemas quality*, due to the central role of multidimensional schemas in OLAP environments. These works often focus on the normalisation of multidimensional schemas [10,11,12] which is only one aspect of their quality. In particular, correct multidimensional schemas should ensure correct summarisation of data at various levels of detail [13]. In [14], a set of metrics for evaluating multidimensional schemas quality is proposed, however the metrics are not related to quality criteria and are specific to ROLAP (Relational OLAP) environments.
- The third category describes global frameworks for data warehouse quality evaluation. The DWQ project (Foundations of Data Warehouse Quality) is representative of this approach [15]. DWQ's framework for data warehouse quality

evaluation distinguishes between the conceptual, logical and physical levels, and defines quality criteria based on these levels and depending on the stakeholders (decision maker, programmer…). DWQ uses and adapts the GQM (Goal-Question-Metric) approach from software quality management.

With respect to previous work, our approach belongs to the second category (multidimensional schemas quality evaluation). We focus not only on multidimensional schemas correctness, but also on other quality criteria of multidimensional schemas. We define metrics which are related to the quality criteria and may be computed (semi-)automatically.

3. The Multidimensional Model

As mentioned in section 1, many multidimensional models have been defined in the literature. However, among all these models, we found no satisfying model encompassing all the important concepts of multidimensional modeling. Therefore, we defined our multidimensional model [4,5], unifying the concepts of the main multidimensional models found in the literature. After an introduction to multidimensional concepts, we present a simplified version of our unified multidimensional model.

3.1. Multidimensional Concepts

Multidimensional models organize data in (hyper)cubes. Therefore, the key multidimensional concepts are *cubes* -which represent *facts* of interest for analysis-, and *dimensions*, i.e. the axes of the cubes.

Fig. 1 shows an example cube, which represents the fact Sale. Sales are analyzed according to three perspectives i.e. dimensions: time, product and geography.

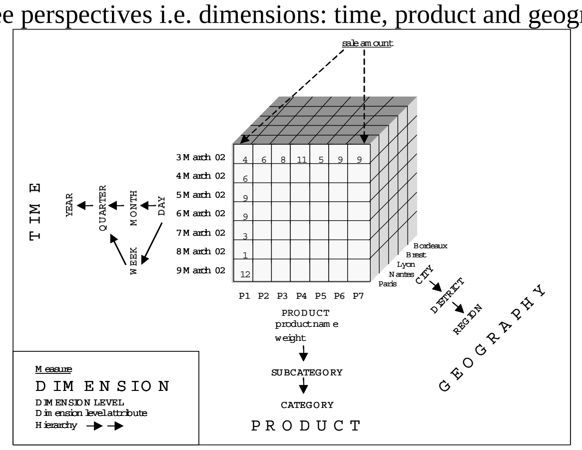

Fig. 1. Multidimensional representation of data

Facts are described by *measures*. A measure, like sale amount (in K€) in **Fig. 1**, is typically a quantitative data. In a cube, the measures correspond to the cells.

Every dimension may consist in one or several aggregation level(s), called *dimension level*s. Dimension levels are organized in *hierarchies*, i.e. aggregation paths between successive dimension levels. In **Fig. 1**, "Day→Week→Quarter→Year" is an example hierarchy. Hierarchies are used in conjunction with aggregation functions (typically, the SUM function) to aggregate ("rollup") or detail ("drill-down") measures. In our example, the sale amount may be totaled at different levels of the Time, Product and Geography dimensions. Often, hierarchies are completed with the special dimension level All, thereby enabling aggregation of measures at the highest possible level (e.g. following the hierarchy "Day→Week→Quarter→Year→All", the total sale amount over the Time dimension may be computed). In addition to being organized in hierarchies, *dimension level*s may be described by *attributes*. For example, the dimension level Product is described by the name and weight of the product. Unlike measures, dimension level attributes are not the object of multidimensional analysis.

Instances of dimension levels are called dimension members. For a given measure in a n-dimensional (hyper)cube, a combination of n dimension members, e.g. (3 March 02, "P1", "Paris"), uniquely identifies a cell and therefore a measure value (4 K€). More specifically, for each axis, the dimension members used as coordinates are instances of the least aggregated dimension level. In the sequel, we will refer to these dimension levels (in our example, Day, Product and City) as "base dimension levels" of their respective dimensions (Time, Product and Geography).

3.2. The Unified Multidimensional Model

Fig. 2 represents our (simplified) multidimensional model with the EER notation.

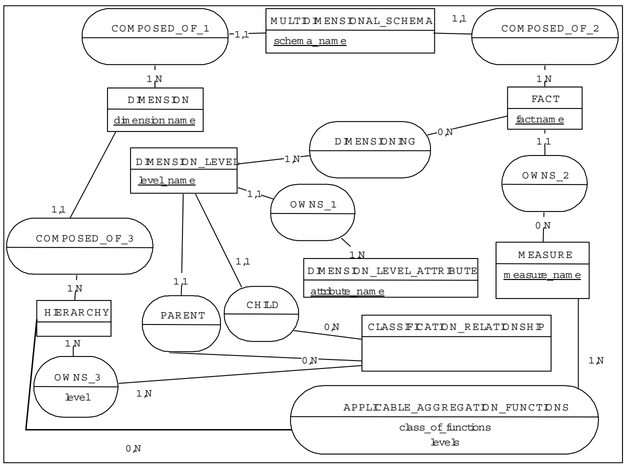

Fig. 2. Unified multidimensional model

Any multidimensional schema is composed of *dimensions* and *facts*, which are interrelated and composed of hierarchies and measures respectively.

Dimensions are defined by grouping *dimension levels* into hierarchies (through classification relationships) and then hierarchies into dimensions.

A *classification relationship* -e.g. "Day→Week"- links a child dimension level to a parent dimension level. Similarly to [16], we define a *hierarchy* -e.g. "Day→Week→Quarter→Year→All"- as a meaningful sequence of classification relationships where the parent dimension level of a classification relationship is also the child of the next classification relationship. In other words, a hierarchy is a meaningful aggregation path between dimension levels. An aggregation path is "meaningful" if valid sequences of drill-down and/or rollup operations can be performed by following the path. Different hierarchies may share common dimension levels and classification relationships. Dimension levels own *dimension level attributes*. Facts are composed of *measures*. Some facts have no measure. Facts are dimensioned by dimension levels (these dimension levels have been called "base dimension levels" in section 3.1). The relationship between a fact and each of its dimensioning dimension levels is called *dimensioning*.

The definition of *applicable aggregation functions* to measures along the different hierarchies is crucial. For every measure, for every dimension level dimensioning the measure (i.e. dimensioning the fact which bears the measure), the set of *aggregation functions* applicable along the different hierarchies starting from the dimension level has to be specified. The unified multidimensional model considers the following functions: SUM, AVG, MIN, MAX, MED (median), VAR (variance), STDDEV and COUNT. Following [17,18,19], we distinguish between three classes of aggregation functions. The first class, which includes all aggregation functions ({SUM, AVG, MIN, MAX, MED, VAR, STDDEV, COUNT}), is applicable to measures that can be summed. The second class ({AVG, MIN, MAX, MED, VAR, STDDEV, COUNT}) applies to measures that can be used for average calculations. The last class contains the single function COUNT. While it is often assumed that the aggregation functions applicable to a measure along a dimension level do not depend on the hierarchies starting from this dimension level, the unified multidimensional model explicitly states that the applicable aggregation functions depend on these hierarchies. For a given hierarchy, applicable aggregation functions may even depend on the levels of the hierarchy i.e. the aggregation functions are applicable only to the first *n levels* of the hierarchy. It may also happen that a measure is not summarisable, whatever the aggregation function and the hierarchy.

We have defined a graphical notation associated with the unified multidimensional model. Dimension levels are represented as 2D rectangles, with their name and attributes. Similarly, facts are represented as 3D rectangles, with their name and measures. Dimensioning relationships are represented as lines and classification relationships are represented with arrows. The graphical multidimensional schema does not include the specification of applicable aggregation functions. This information is specified in a separate table.

4. Multidimensional Schemas Quality Evaluation Framework

The objective of this paper is to present a systematic way to quantitatively evaluate the quality of multidimensional schemas.

We consider the quality of multidimensional schemas according to three views :

- the specification view concerned with the data warehouse designer's objectives,
- the usage view dealing with the decision maker's requirements,
- the implementation view related to the developer's concerns.

For the sake of this paper, we focus on the **specification** view, which captures the degree of fitness of the multidimensional schema with reality and more specifically with the user's needs. We have identified the following criteria: *legibility*, *expressiveness*, *simplicity* and *correctness*.

4.1. Legibility

The legibility (or readability) expresses the ease with which a multidimensional schema can be read. To measure legibility, we propose two subcriteria, namely *minimality* and *zoom in – zoom out facility*.

A schema is minimal when every aspect of the requirements appears only once [20]. To measure minimality, we propose to define two subcriteria : *non-redundancy and factorisation degree*. For *non-redundancy*, we propose the following metric :

Non-redundancy= $\dfrac{\sum_{S} wi\ NB(Ci) - wi\ NB_R(Ci)}{\sum_{S} wi\ NB(Ci)}$	Where Ci belongs to {fact, dimension, dimension level}. NB(Ci) calculates the number of elements of type Ci, $NB_R(Ci)$ calculates redundant elements of type Ci in the current schema S, w_i is the weight associated with Ci.

To illustrate this metric, let's consider the example schemas of **Fig. 3**.

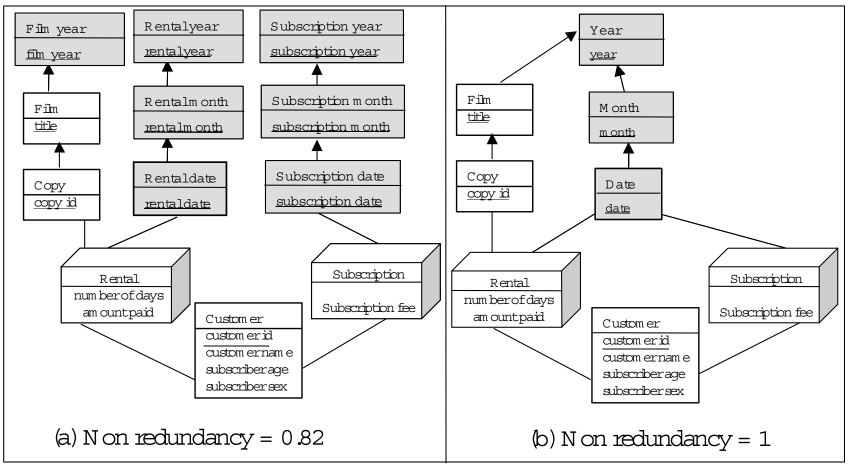

Fig. 3. Measuring non-redundancy

Schemas (a) and (b) from **Fig. 3** represent the same reality. However schema (a) makes a distinction between rental, subscription and film dates whereas schema (b) manages the three dates using the same concepts.

According to the non-redundancy quality criterion, schema (b) is less redundant and the non-redundancy value of schema (b) is higher that the one of schema (a).

The second sub-criterion for minimality is the factorisation degree quality criterion. The metric associated with this criterion is the following:

Factorisation degree = $\sum_{H}\left[\sum_{Ci}\left(1-\frac{DEF(Ci)}{USE(Ci)}\right)\Big/NB(Ci)\right]\Big/NB(H)$	Where H is a hierarchy of dimension levels, Ci Є {dimension level attribute}, DEF(Ci) counts the number of occurrences of an element Ci in a hierarchy, USE(Ci) counts the number of facts in the multidimensional schema using this dimension level attribute, NB(H) is the number of hierarchies in the schema, and NB(Ci) the total number of dimension level attributes.

Applying the factorisation degree metric on schemas (a) and (b) depicted in **Fig. 3** produces the following results:

Schema	(a)	(b)
Factorisation degree	0.25	0.58

The degree of factorisation increases when each hierarchy of dimension levels is related to several facts simultaneously. In this case, each attribute from the dimension levels composing the hierarchy is defined once and used a number of times corresponding to the number of facts related to the hierarchy.

The second sub-criterion for legibility is the zoom in - zoom out facility. This criterion measures the possibility of viewing a schema at several levels of granularity. To measure this criterion, we propose the following metric:

Zoom in zoom out facility $= \sum_{1}^{NB(F)} 1 - 1/ZL(Fi) \Big/ NB(F)$	Fi is a fact from the schema, NB(F) is the number of facts and ZL(Fi) is the maximal depth of the hierarchies of dimension levels of the fact Fi.

Let's consider the fact named "Rental" from **Fig. 3**. This fact can be viewed at 4 levels of detail (**Fig. 4**):

- At the first level, "Rental" is seen as a fact representing the rental of a film copy by a customer at a given date.
- At the second level, the schema allows the visualisation of rentals by film, by copy medium, by customer type, by month or by subscriber city.
- At the third level, rentals could be viewed by film type, by country or by year and finally,
- At the fourth level, rentals could be viewed by film continent.

Applying the metric on this schema calculates a value of zoom in - zoom out facility equal to 0.75.

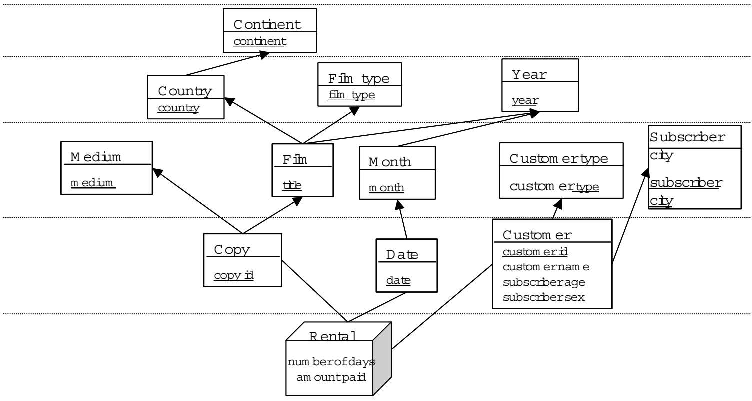

Fig. 4. Zoom in-zoom out facilities

4.2. Expressiveness

A schema is said to be expressive when it represents users requirements in a natural way and can be easily understood without additional explanation [20]. In the case of multidimensional schemas, expressiveness measures the variety of analyses that the decision makers will be able to perform based on the schema. We distinguish between two levels of *expressiveness*, namely *fact* and *schema expressiveness*.

Measuring Fact Expressiveness

We assume that fact expressiveness depends on:

- the number of measures describing the fact,
- the number of dimensions i.e. the number of base dimension levels dimensioning the fact,
- the number of dimension levels related to the fact,
- and the aggregation functions that can be applied on the measures of the fact.

To take these aspects into account we identified four sub-criteria for fact expressiveness, namely: *Fact richness*, *Fact dimensioning*, *Fact analysability* and *Fact summarisability.*

Fact richness : The underlying assumption for this metric is that the richness of a fact F depends on the calculation potential captured by its measures. This potential could be calculated locally if taking into account only the schema to which F belongs (*Local Fact Richness*) or globally with regard to a set of alternative multidimensional schemas (*Global Fact Richness*). The metrics are the following:

Local Fact Richness(F) $= \dfrac{\text{NBmeasures (F)}}{\text{NBmeasures (S)}}$	Where F is a fact from a multidimensional schema S. Nbmeasures is a function counting the number of measures contained in either a fact or a schema.

Note that local fact richness enables only the comparison of calculation potential of facts within the same schema. This measure is not relevant when the concern is to compare several alternative multidimensional schemas representing the same reality. In this situation a global fact richness is more suitable:

Global Fact Richness(F) $= \dfrac{\text{NBmeasures(F)}}{\text{NBmeasures}(\bigcup_{1}^{N}(\text{Si}))}$	Where F is a fact from a multidimensional schema S. NBmeasures counts the number of measures contained in either a fact or a schema and $U(S_i)$ calculates a syntactic union schema from a set of alternative schemas

Fact dimensioning : This criterion assumes that a n-dimensioned fact is more expressive than a m-dimensioned fact when n>m. Similarly to the fact richness evaluation, we distinguish between local and a global fact dimensioning.

Local Fact Dimensioning(F) $= \dfrac{\text{NBDimensions(F)}}{\text{MAX(NBDimensions(Fi) Fi} \in \text{S}}$	Where F is a fact from a multidimensional schema S. NbDimensions is a function counting the number of base dimension levels of the fact F. Max is a function calculating the maximal value among a set of values.

For the global fact dimensioning metric, we consider a set of alternative multidimensional schemas representing the same reality. The metric is the following:

Global Fact Dimensioning(F) $= \dfrac{\text{NbDimensions(F)}}{\text{MAX(NbDimensions(Fi)Fi} \in \bigcup_{1}^{N}\text{(Si)}}$	Where F is a fact from a multidimensional schema S. NbDimensions is a function counting the number of base dimension levels dimensioning the fact F. Max is a function calculating the maximal value among a set of values. $U(S_i)$ is a function calculating the set of alternative schemas $(S_1, \ldots S_N)$.

Fact analysability : This criterion refines the one concerning fact dimensioning. Indeed, the fact dimensioning criterion takes into account only the base dimension levels related to the fact. However, a fact can be analysed based on the base dimension levels and all the dimension levels related to these base dimension levels. Let's consider the fact "Rental" from **Fig. 3**-(a). According to the fact dimensioning criterion, only the dimension levels {Copy, Customer, Rental date} are taken into account. For fact analysability, we will consider also the other dimension levels allowing the analysis of the "Rental" fact, following the hierarchies related to this fact. The set of these dimension levels for the "Rental" fact is {Copy, Customer, Rental date, Medium, Film, Film type, Film year, Country, Continent, Customer type, Subscriber city, Rental month, Rental year}, as all of the measures defined in the "Rental" fact could be analysed according to each of these dimension levels. For the same quality measurement considerations taken into account for fact dimensioning and fact richness, we have a local and a global fact analysability metrics described below:

Local Fact Analysability(F) $= \dfrac{\text{NbADimensions(F)}}{\text{MAX(NbADimensions(Fi) Fi} \in \text{S)}}$	Where F is a fact from a multidimensional schema S. NbADimensions is a function counting the number of dimension levels related to a fact F.

For the global fact analysability metric, we consider a union schema in which all the facts appearing in the alternative schemas are represented and each fact is related to a maximal number of dimension levels deduced from the several schemas. Such a union schema is not always correct because the union here is purely syntactic and does not correspond to an integration of the schemas.

Global Fact Analysability(F) = $\dfrac{NbADimensions(F)}{MAX(NbADimensions(Fi)\ Fi \in \bigcup_{1}^{N}(Si))}$	Where F is a fact from a schema S. NbADimensions is a function counting the number of dimension levels related to a fact F. $U(S_i)$ is a function calculating the set of alternative schemas ($S_1, \ldots S_N$).

Fact summarisability : this criterion is related to the applicability of aggregation functions on the measures of a given fact. Let's consider again the example from **Fig. 3**-(a). For the fact "Rental", all aggregation functions (SUM, AVG, COUNT, etc...) are applicable to the measure "amount paid". Moreover, these functions are applicable at each of the dimension levels related to the fact "Rental". However, the "number of days" of a rental may not be summed along the dimension level Copy (e.g. if a customer has rented two copies simultaneously, the duration of his rental is not the sum of the duration of rental of the two copies); therefore, in this case an aggregation function like AVG or MAX may be used. To evaluate fact summarisability, we associate values to the classes of aggregation functions presented in section 3 ({SUM, AVG, MIN, MAX, MED, VAR, STDDEV, COUNT} has the highest value, {AVG, MIN, MAX, MED, VAR, STDDEV, COUNT} has a lower value…). The metric proposed for fact summarisability is based on both the class of functions that can be applied and the number of dimension levels for which this application makes sense.

Local Fact Summarisability (F_k) = $\dfrac{\sum_j \sum_i FuncApp(DLi, Mjk)}{\sum_k \sum_j \sum_i FuncAppl(DLi, Mjk)}$	Where F_k is a fact from a multidimensional schema S. FuncApp(DL_i,Mj) is a function associating a value for the aggregation functions applicability for the measure M_{jk} and the dimension level DL_i. M_{jk} is a measure belonging to the fact F_k and DL_i is a dimension level related to F_k.

For the global fact summarisability metric, we consider again a union schema and the metric is the following:

Global Fact Summarisability (F_k) $= \dfrac{\sum_j \sum_i FuncApp(DLi, Mjk)}{\sum_l \sum_j \sum_i FuncAppl(DLi, Mjl)}$	Where F_k is a fact from a multidimensional schema S. FuncApp(DL_i,Mj) is a function associating a value for the aggregation functions applicability for the measure M_{jk} and the dimension level DL_i. M_{jk} is a measure belonging to the fact F_k and DL_i is a dimension level related to F_k. l is the number of facts from the union schema.

Measuring Schema Expressiveness

For schema expressiveness, we suggest an average calculated for each of the fact expressiveness sub-criteria. We will not detail the metrics due to space limitations.

Simplicity

A schema is said to be simple if it contains the minimum possible constructs. Our measure of simplicity is based on the assumption that the complexity of a multidimensional schema grows with the number of concepts (including dimensioning relationships and classification relationships).

Simplicity (S)= $\frac{NB(F)+NB(DL)}{NB(F)+NB(DL)+NB(link)}$	Where NB(F), NB(DL) correspond respectively to the number of facts and dimension levels in a schema S. NB(link) is the number of links (dimensioning relationships and classification relationships) in S.

Correctness

Correctness is used in a wide range of contexts leading to very different interpretations. A schema is syntactically correct when concepts are properly defined in the schema [20]. To measure correctness, we suggest the following metric:

Correctness(S) = $\frac{\sum_{i=1}^{N}(VERIF(C_i) - ERR(C_i))}{\sum_{i=1}^{N}(VERIF(C_i))}$	Where VERIF() is a function calculating the number of characteristics to be verified on an element Ci of the current multidimensional schema. This number is the same for all the occurrences of the same type concept. ERR() is a function calculating the number of errors depicted on an element Ci. N is the number of elements in a schema S

5. Conclusion and Further Research

This paper proposed an approach for multidimensional schemas quality assessment, based on a quality evaluation framework with three complementary viewpoints. We focused on the specification view and presented criteria and associated metrics, which can be computed (semi-)automatically. Our approach is especially useful when the data warehouse designer has to choose between several alternative designs.

Further work includes the empirical/theoretical validation and refinement of the metrics proposed in this paper, as well as the investigation of the usage and implementation views. We are currently working on these issues.

6. Acknowledgements

We thank colleagues from CEDRIC-CNAM, more specifically Jacky AKOKA and Isabelle COMYN-WATTIAU, for their helpful comments on this paper. This work is partly funded by the Research Center of ESSEC.

7. References

1. Chen, P.P.: The entity-relationship model – toward a unified view of data. ACM TODS, volume 1, number 1, March 1976
2. Blaschka, M., Sapia, C., Höfling, G., Dinter, B.: Finding your way through multidimensional data models. DEXA Workshop on Data Warehouse Design and OLAP Technology (DWDOT '98), Vienna, Austria, 1998
3. Vassiliadis, P., Sellis, T.: A survey of logical models for OLAP databases. SIGMOD Record, volume 28, number 4, December 1999
4. Akoka, J., Comyn-Wattiau, I., Prat, N.: Dimension hierarchies design from UML generalizations and aggregations. 20th International Conference on Conceptual Modeling (ER 2001), Yokohama, Japan, November 2001
5. Prat, N., Akoka, J., Wattiau, I.: A data warehouse design method based on UML, to be submitted for publication.
6. Si-Saïd, S., Akoka, J., Comyn-Wattiau, I.: Conceptual Modeling Quality – From EER to UML Schemas Evaluation. Proceedings of the 21th International Conference on Conceptual Modeling (ER2002), LNCS, Springer.
7. Si-Saïd, S., Akoka, J., Comyn-Wattiau, I.: Measuring UML Conceptual Modeling Quality – Method and Implementation. Proceedings of the BDA Conference, Ed. P. Pucheral, Collection INT, France, 2002
8. Hufford, D.: Data warehouse quality: special feature from January 1996. DM Review, January 1996
9. Ballou, D.P., Tayi, G.K.: Enhancing data quality in data warehouse environments. Communications of the ACM, volume 42, number 1, January 1999
10. Lechtenbörger, J., Vossen, G.: Multidimensional normal forms for data warehouse design. Information Systems, volume 28, number 5, 2003
11. Lehner, W., Albrecht, J., Wedekind, H.: Normal forms for multidimensional databases. 10th International Conference on Statistical and Scientific Database Management (SSDBM '98), Capri, Italy, July 1998
12. Levene, M., Loizou, G.: Why is the snowflake schema a good candidate for data warehouse design? Information Systems, volume 28, number 3, 2003
13. Lenz, H.-J., Shoshani, A.: Summarizability in OLAP and statistical data bases. 9th International Conference on Statistical and Scientific Database Management (SSDBM '97), Olympia, Washington, USA, August 1997
14. Calero, C., Piattini, M., Pascual, C., Serrano, M.A.: Towards data warehouse quality metrics. 3rd International Workshop on Design and Management of Data Warehouses (DMDW 2001), Interlaken, Switzerland, June 2001
15. Jarke, M., Jeusfeld, M., Quix, C., Vassiliadis, P.: Architecture and quality in data warehouses: an extended repository approach. Information Systems, volume 24, number 3, 1999
16. Tsois, A., Karayannidis, N., Sellis, T.: MAC: conceptual data modeling for OLAP. 3rd International Workshop on Design and Management of Data Warehouses (DMDW 2001), Interlaken, Switzerland, June 2001
17. Lehner, W.: Modeling large scale OLAP scenarios. 6th International Workshop on Extending Database Technology (EDBT '98), Valencia, Spain, March 1998
18. Pedersen, T.B., Jensen, C.S.: Multidimensional data modeling for complex data. 15th International Conference on Data Engineering (ICDE '99), Sydney, Australia, March 1999
19. Rafanelli, M., Ricci, F.: Proposal of a logical model for statistical databases. 2nd International Workshop on Statistical Database Management (SSDBM '83), Los Altos, California, September 1983
20. Batini, C., Ceri, S., Navathe, S.B.: Conceptual database design : An entity relationship approach. Benjamen Cummings, Redwood City, California, 1992