

# Building Conflict-Aware Profiling Ontology from Data Warehouses

Eya Ben Ahmed  
University of Tunis  
Higher Institute of  
Management,  
Tunisia

Ahlem Nabli  
Sfax University  
Faculty of Science  
of Sfax,  
Tunisia

Faiez Gargouri  
Sfax University  
Higher Institute of Computer  
and Multimedia of Sfax  
Tunisia

## ABSTRACT

User profiles or user models are crucial in many areas in which it is essential to obtain knowledge about users of software applications such as data warehouse technologies. To enhance the personalized services, group profiles are derived through combining individual user profiles in order to represent group modelling. In this paper, we propose a new representation of group profiles in OLAP context using the ontological modelling. Our main aim is to semantically enrich the representation of group preferences in data warehouses. Our ontology is validated using set of real collected OLAP query logs in stock market area.

## General Terms

Profiling, Data warehouses, Ontology.

## Keywords

Profiling, groupization, preferences, data warehouse, profiling ontology, modular ontology.

## 1. INTRODUCTION

The user profile is a description of the most relevant facts of user known as preferences and modeling particular interests of user in an application field. The motivation behind building user profiles is that users differ in their preferences, interests and goals when using applications. Discovering these differences is vital for providing users with personalized services. Such personalized services were somewhat effective in the beginning, but they gradually declined with accuracy over time because the amount of stored user information has been amplified. However, it is useful to provide user-oriented services based on preferences.

In many application domains, several users handle the system. Therefore, drawing all their users profiles will be a laborious task. Nevertheless, gathering users in respect to their common interests to describe their similar preferences may be an efficient solution for this complex issue. Compared to individual user profiles, group profiles combine individual user profiles to represent group modeling. Gathered profiles represent group of members' preferences instead of modeling each user preferences.

They are crucial in several application fields where it is required to provide recommendations to groups of users rather than to individual users. Typical examples of such fields are tourism recommendation systems (i.e. INTRIGUE system [1] which based on group characteristics recommends places to visit for tourist groups) and TV program recommendation systems where combination of individual user profiles is used in an adaptive television application [16].

Among the application domains, we stress on data warehouse which collected transactional data coming from several

sources, filtered, merged and stored them in a central repository. The content of a data warehouse is interactively analyzed using the OnLine Analytical Processing (OLAP) in order to support a broad range of knowledge seeker's tasks to take the relevant decision at the appropriate time.

Drawing decision marker profiles from data warehouses attracts a strong interest in data warehousing community during the last few years [2, 3, 6, 7]. The discovered profiles sum up preferences closely related to the most occurring subschema in performed OLAP analysis.

Despite the investigated efforts, problems still arise when dealing with the representation of analyst preferences, their semantic issues and their exploration to precisely define the decision marker behavior.

Actually, the ontology domain provides several solutions modeling the knowledge engineering [9]. Indeed, the ontology is an explicit specification of a conceptualization [8]. It models the representation of semantic information offering definitions and descriptions of the concepts of a target application domain [12, 13, and 15].

Recently, we assist to spectacular explosion of methodologies for development of ontologies and their efficient use in various domains, among others decision support systems [11].

In this work, we might be able to find another way of tackling the problem of analyst preferences expressing: we propose an ontological representation of groups of analysts' preferences in data warehouse context. Indeed, the process of uniting the users into groups to better meet their expectations during the use of system is called 'groupization'. The promising challenge of performing groupization is the best detection of such groups using identification criteria.

In our previous work [4], the groupization was performed based on analysts past activities using four defined criteria, namely, *the function*: assuming that analysts working at the same position have similar preferences, *the responsibilities* to accomplish the goals: the analysts have communally goals that may be operational for short-term, tactical for medium-term or strategic for long-term, *the source of group identification* which may be explicit or implicitly learned, and *the dynamicity of the identified groups* which can be updated dynamically or remained static. We demonstrated that the function criterion is the most discriminating for analysts groups identification. The derived groups may gather users usually having the same interests. However, in some cases, each analyst may manipulate differently the data warehouse in respect to his current expectations. Accordingly, the concept of conflict in the same group was handled in [3]. According to this aspect, the discovered preferences are categorized in two distinct classes [3]: (i) consensual preferences referring to same MDX queries launched by all similar function analysts;

(ii) conflicting preferences related to disjoint MDX queries for the same function analysts.

In fact, we introduce a new conflict-aware profiling ontology including modular component vocabularies to represent group interests and model similar behavior. Our innovative approach is motivated by several key motivations behind our proposal : (i) the representation of the semantic relations between entities; (ii) the formalization of the group preferences avoiding any plausible ambiguity and inconsistency already present in the classical preferences modeling; (iii) facilitating the preferences dissemination and reuse; (iv) the improvement of the communication between the users of the ontology thanks to the particular sense of the vocabulary and meaning of the data warehouse domain terms.

The remainder of this paper is structured as follows. After summarizing the related work in section 2, we present, in section 3, our conflict-aware profiling ontology. Section 4 details the profiling module of our innovative ontology. Section 5 describes the multidimensional module. Our case study is sketched in section 6. We conclude, in section 7, the paper and outline our future work.

## 2. RELATED WORK

Several approaches to analyst profiling were devised in the data warehousing context. We can distinguish two trends of research works dedicated to profile construction method:

(i) The first category implicitly integrates the user profile as dynamic constraints involved during the query process [6].

(ii) The second category expresses the knowledge seeker preferences using explicit behaviour modelling through personalized MDB schema [7] or an extension of MDX language as preference algebra [2].

The comparison of the pioneering profiling approaches in data warehousing, shown by table 1, reveals an exclusive exploration of individual analysts' profiles which are semantically poor and overlook the conflict of interest embodied in the generated sub-schema during OLAP analysis.

Slightly different from these approaches addressing the issue of analyst profiling, our key idea consists in representing the preferences in form of domain ontology. In previous works [3, 4], we performed groupization in data warehouses based on function-criterion and we dealt with conflict-aware profiles. In this paper, we extend our previous proposals and introduce an original conflict-aware profiling ontology.

To the best of our knowledge, our work is the first that proposes to represent the decision maker preferences in form of semantic ontological concepts and handle the conflict aspect in decision makers preferences.

**Table 1. Comparison of profiling approaches in OLAP context.**

Construction Preferences Method		Granularity		Modelling				Conflict	
		Individual	Group	Personalized MDB schema	Preference algebra	Dynamic constraints	Ontology	Consensual	Conflicting
Implicit	Golfarelli et al. [6]	X			X			X	
	Bellatreche et al. [2]	X		X				X	
	Jerbi et al. [7]	X				X		X	
	<b>Our work</b>		X				X	X	X

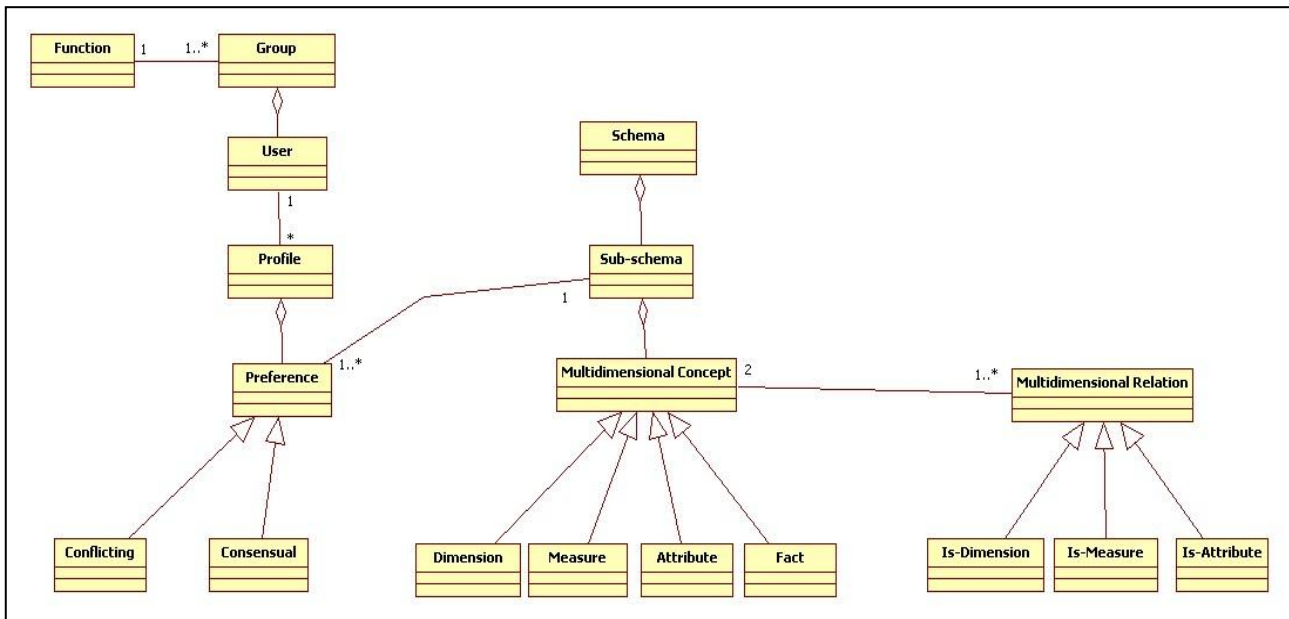
## 3. CONFLICT-AWARE PROFILING ONTOLOGY FROM DATA WAREHOUSES

A user profile is a representation of information about an individual user that is essential for our application. Such a derived profile is crucial in various in which it is essential to obtain knowledge about users of software applications.

Learning those user preferences in intelligent profiles is closely dependent on the number of analysts. Indeed, the growing number of analysts may prevent this automatic discovery of preferences. To overcome this limitation in the data warehouse area, we expand the personalization process through combining similar individual's data analysts to improve the process of preferences learning.

This extension of this mechanism through considering data related to all group members instead of data associated to single user is called "groupization" [4]. It is well-recognized that several criteria are more useful than others for group detection based on their past activities. Our findings, in [4,5], demonstrate that groupization improves upon personalization for several group types, mainly for the groupization based on held function.

In collaborative context, the gap between the analysts behavior may be more extensive than in individual context. Thus, we integrate the conflict notion in analysts preferences. In [3], we argued that preferences may be categorized into two main pools: (i) consensual preferences referring to common preferences (*i.e.*, similar launched MDX queries) for the same function decision makers; (ii) conflicting preferences related to divergent preferences (*i.e.*, diverse run MDX queries) for similar function analysts.



**Fig 1 : The meta-model of conflict-aware profiling ontology in data warehouse**

In this paper, we aim to palliate several drawbacks characterizing the representation of analyst profile, namely, (i) avoiding any ambiguity on expressing preferences due to the lack to data warehouse profiling standard; (ii) assisting the preferences reuse in several cases, namely in profiling update after any maintenance of the data warehouse schema or any human–computer interface personalization; (iii) describing relevant meaning of preferences terms in order to improve of communication between the users of the ontology for OLAP services personalization.

Thus, we introduce an innovative modeling of analysts preferences using ontology background. Indeed, the conflict-aware profiling ontology is a representation of knowledge related to group of analysts preferences in decisional context. It can be defined as a referential of multidimensional concepts of a field, and their semantic and multidimensional relations. Figure 1 illustrates the meta-model of our conflict-aware profiling ontology.

Motivated by the issue that modular ontologies ensure an efficient selective reuse of the component ontologies [14], we choose to build our profiling ontology in a modular manner. First, we develop small modules. Then we compose them to form our modular ontology. Such obtained profiling ontology includes a set of semantically related component ontologies, fundamentally summarized as follows:

**Profiling module** depicting the conceptualization of profiling concepts and their associated relations;

**Multidimensional module** describing the conceptualization of decisional classes and their relations.

Such modular design of our ontology is depicted by figure 2.

#### 4. PROFILING MODULE

The decision makers' preferences are represented in ontological format to enhance their semantic description. The main semantic concepts are described in the following.

**User concept** is the decision maker. It is represented by the user concept (Usr).

*Example 1.* A typical example of user concept is Jean Lewandowski.

**Group concept** is composed of users. It is represented by the group concept (Grp).

*Example 2.* An example of group concept is the cluster of Jean Lewandowski and Mark Geneve.

**Function concept** is related to the hold function of the group. It is represented by the function concept (Fun).

*Example 3.* Portfolio manager is an example of function concept.

**Profile concept** consists in the profile of the group. It is represented by the profile concept (Prf).

*Example 4.* An example of profile concept is depicted by figure 3.

**Preference concept** is the elementary component of the profile. It is represented by the preference concept (Pre).

*Example 5.* An example of preference concept is stock market speculation.

**Conflicting concept** represents the disjoint preference. It is represented by the conflicting concept (Cof).

*Example 6.* 2012 is an example of conflicting concept.

**Consensual concept** is related to similar preference. It is represented by the consensual concept (Cos).

*Example 7.* A typical example of consensual concept is stock exchange index.

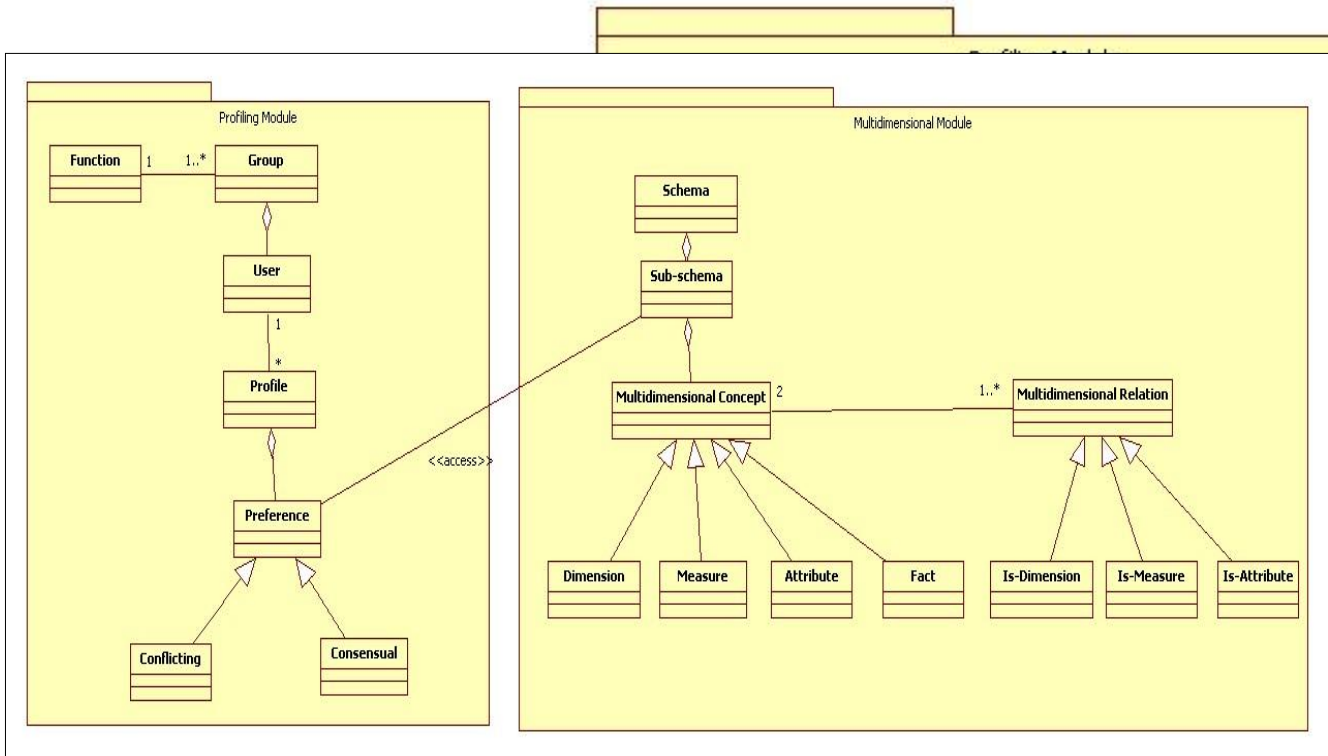


Fig 2 : The meta-model of modular design of our conflict-aware profiling ontology.

<i>Portfolio manager group profile</i>
(Stock market exchange index) : consensual
(Time: Year: 2012) : consensual
(Time: Month: May) : conflicting
(Title: Sector: Health) : conflicting
....

Fig 3 : Example of investor group profile.

## 5. MULTIDIMENSIONAL MODULE

Focusing in collecting the analyst preferences, we stress on the multidimensional aspect of our ontology. Interestingly, we describe the multidimensional concepts and their semantic relations.

### 5.1 Concepts

**Schema concept** is the form of data warehouse multidimensional model. Such schema can exist as a star schema, a snowflake schema, or a fact constellation schema. It is represented by the schema concept (Sch);

*Example 8.* An example of schema concept is stock market speculation schema sketched by the figure 4.

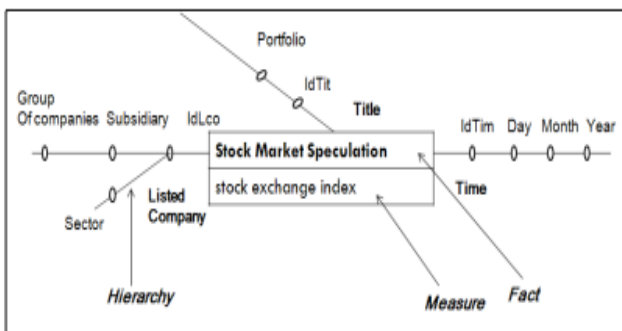


Fig 4 : Stock market fact schema.

**Sub-schema concept** is a part of the schema concept. It is represented by the sub-schema concept (SSc);

*Example 9.* An example of sub-schema concept is illustrated by figure 5.

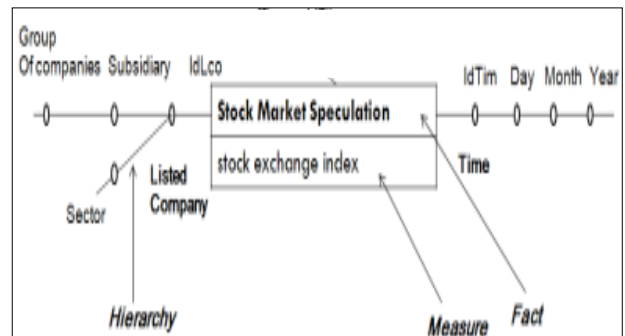


Fig 5 : Stock market fact sub-schema.

**Multidimensional concept** is the elementary component of the sub-schema. It is represented by the multidimensional concept (Mld);

*Example 10.* An example of multidimensional concept is the listed company dimension.

**Fact concept** is the subject of analysis and it always represented by frequently updated data . It is represented by the fact concept (Fac);

*Example 11.* An example of fact concept is stock market speculation.

**Dimension concept** The dimension corresponds to perspective or entity with respect to which measures are analyzed. It is represented by the dimension concept (Dim).

*Example 12.* An example of dimension concept is time dimension.

**Attribute concept** A dimension is made up of attributes. It is represented by attribute concept (Atr).

**Example 13.** 2012 is an example of attribute concept.

**Measure concept** A numerical property of a fact. It is represented by measure concept (Mea).

**Example 14.** An example of measure concept is stock exchange index.

## 5.2 Relations

We outline the identified relations in this subsection.

**Is-Dimension** ( $C_i, C_j$ ) the semantic type of the concept  $C_i$  is Fac, the semantic type of the concept  $C_j$  is Dim and  $C_j$  is a dimension for the fact  $C_i$ ;

**Is-Measure** ( $C_i, C_j$ ) the semantic type of the concept  $C_i$  is Fac, the semantic type of the concept  $C_j$  is Mea and  $C_j$  is a measure for  $C_i$ ;

**Is-Attribute** ( $C_i, C_j$ ) the semantic type of the concept  $C_i$  is Dim, the semantic type of  $C_j$  is Atr and  $C_j$  belongs to  $C_i$ .

## 6. CASE STUDY

It was worth the effort to experience in practice the potential benefits of our proposed ontology.

First, we introduce our framework then we illustrate our proposal with a financial example drawn from such a created data warehouse.

### 6.1 Financial Data Warehousing Framework

We built a data warehouse in stock market area. Indeed, stock exchanges are corporations in which public institutions and companies can raise long-term capital for development projects and/or for businesses in order to assist growth. They offer trading facilities for investors to buy and sell securities. A *stock market index* is used to check the performance of a group of stocks. It is a statistical metric reflecting the composite value of its components and a measure of the return that would accrue to the holder of a particular set of equities.

Our designed data warehouse aims to analyze the stock average and assess the performance of the stock market. Its major goal is to investigate the stock indexes fluctuations for performance estimation of the stock market. The logical design of our constructed data warehouse is depicted by Figure 4.

### 6.2 Ontology Building and Results

As a financial data warehouse, it enables several economic actors such as investors and managers to analyze relevant multidimensional data using OLAP technology. Such conducted analyses using MDX queries are stored in OLAP query logs. An example of OLAP query log is shown by figure 6.

```
<?xml version="1.0" encoding="UTF-8"?>
  <Log session = " portfolio manager ">
    <Query id = "X">
      SELECT [Measures].[stock exchange index] ON
      COLUMNS, [Time].[All] ON ROWS
      FROM Stock Market Speculation
      WHERE ( [Title].[Portfolio] )
    </Query>
  </Log session>
</xml>
```

**Fig 6 : Extract of OLAP query log of portfolio manager.**

We gathered the logs in clusters according to the analyst function criterion because our findings in [4] prove that decision makers working at the same position have similar run queries on designed data warehouse.

Figure 7 illustrates the functional architecture of our case study. Based on derived clusters, we build our conflict-aware profiling ontology. Its goal is to precisely describe the preferences of each group of analysts performing the same function. Such designed ontology defines not only the basic terms and relations comprising the vocabulary of a data warehouse area, but also the rules for combining terms and relations to define extensions to the warehousing vocabulary.

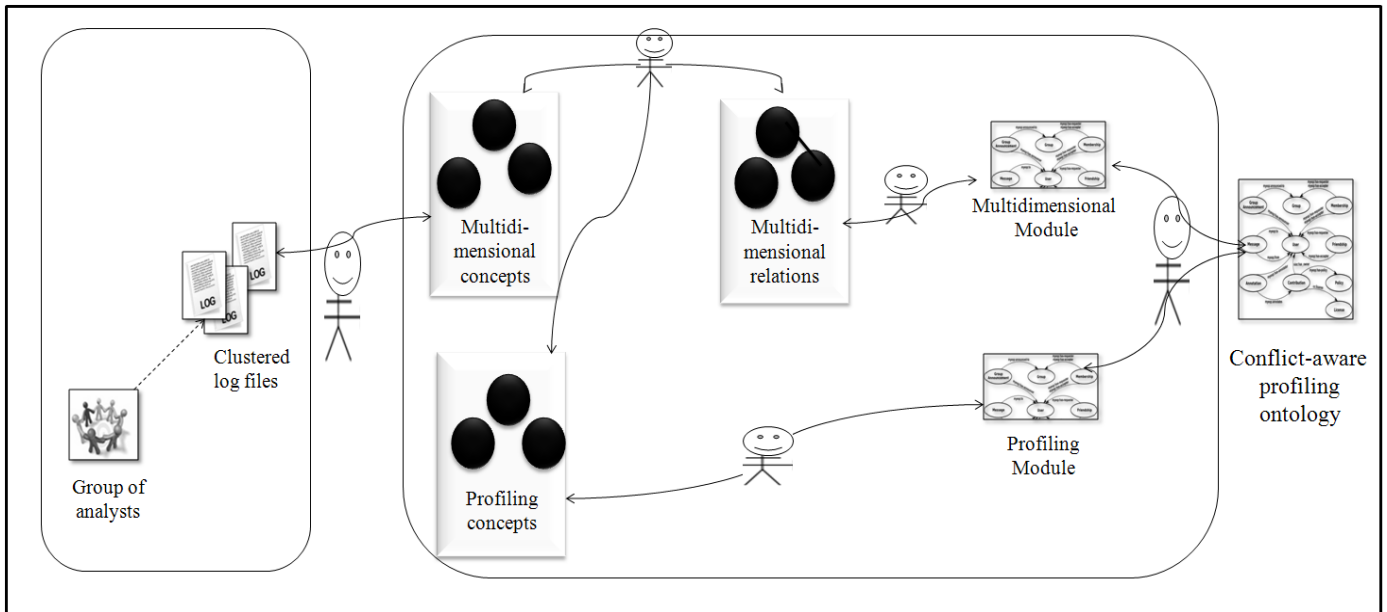
To build our ontology, we followed those steps. First, we identified our key concepts and relationships in our application domain, such as *user, group, function, sub-schema, dimension, fact, attribute.....etc* and the relations between them, namely, *is-fact, is-dimension, is-attribute...etc*.

Then, we produced precise definitions for such concepts and relationships as already presented in the previous sections. For example, the **function concept** is related to the hold job by set of users.

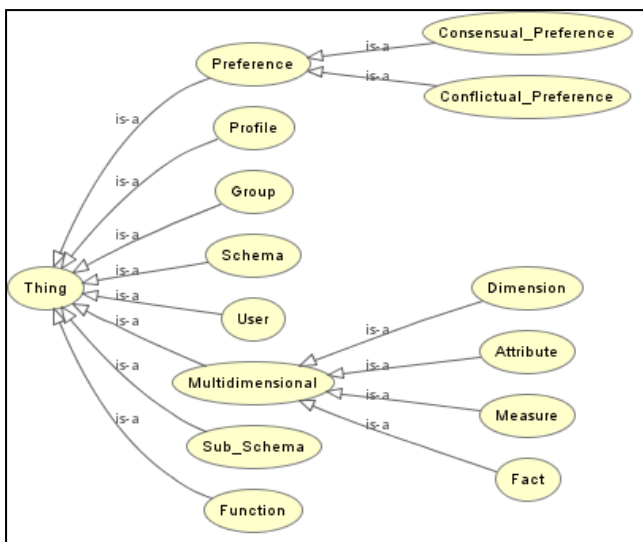
After that, we identified the terms referring to such concepts and relationships manually extracted from OLAP query logs. An illustrative example of term referring to function concept is **portfolio manager**.

During the building process, the knowledge engineer must distinguish between ontological profiling module and ontological multidimensional components.

At the end of this stage, we provided an explicit representation of the concepts using the OWL representation language. Thus, our original ontology has been implemented in OWL, since it has been modelled with the Protégé tool and the OWL plug-in [10]. Figure 8 depicts a part of our designed conflict-aware profiling ontology.



**Fig 7 : Functional architecture of our case study.**



**Fig 8 : Fragment of the conflict-aware profiling ontology**

## 7. CONCLUSION

In this paper, we presented an innovative conflict-aware profiling ontology. Two main modules are sketched: (i) *Profiling module* for defining the group preferences ; (ii) *Multidimensional module* for describing the conceptualization of decisional classes and their relations. Our main aim is to semantically enrich the representation of group preferences in data warehouses.

This work led to several interesting research directions. Among them, we can explore: (i) the integration of different data sources to enrich the preferences of groups of analysts; (ii) using the created ontology to personalize the stock market OLAP technology in order to provide suitable information to

Financial decision maker; (iii) handling incremental evolution of our constructed profiling ontology and (iv) semantic annotating our ontology according to analysts feedbacks.

## 8. REFERENCES

- [1] Ardissono, L., Goy, A., Petrone, G., Segnan, M., Torasso, P.: Ubiquitous User Assistance in a Tourist Information Server. *International Conference on Adaptive Hypermedia and Adaptive Web Based Systems*. LNCS, Springer, Heidelberg, 14-23 (2002)
- [2] Bellatreche, L., Giacometti, A., Marcel, P., Mouloudi, H., Laurent, D.: A personalization framework for OLAP queries. *Proceedings of the 8th ACM international workshop on Data warehousing and OLAP*, Bremen, Germany, 9-18 (2005)
- [3] Ben Ahmed, E., Nabli, A., Gargouri, F.: Building MultiView Analyst Profile From Multidimensional Query Logs: From Consensual to Conflicting Preferences. *The International Journal of Computer Science Issues (IJCSI)*, January 2012, Volume 9, Number 1 (2012) 124-131
- [4] Ben Ahmed, E., Nabli, A., Gargouri, F.: Performing Groupization in Data Warehouses: Which Discriminating Criterion to Select?. *Proceedings of the 17th International conference on Applications of Natural Language Processing to Databases (NLDB'12)*, June 2012, Springer-Verlag (2012) 234-240
- [5] Ben Ahmed, E., Nabli, A., Gargouri, F. : SHACUN : Semi-Supervised Hierarchical Active Clustering Based on Ranking Constraints, *Proceedings of the 12<sup>th</sup> Industrial Conference on Data Mining (ICDM'12)*, July 2012, Springer-Verlag (2012) 194-208
- [6] Golfarelli, M., Rizzi, S., Biondi, P.: myOLAP: An approach to express and evaluate OLAP preferences. *IEEE Transactions on Knowledge and Data Engineering* ( 2011) 1050-1064

- [7] Jerbi, H., Ravat, F., Teste, O., Zurfluh, G.: Management of context-aware preferences in multidimensional databases, *IEEE International Conference on Digital Information Management*, London, UK, pp. 669-675 (2008)
- [8] Gruber, T.R.: Toward Principles for the Design of Ontologies. *Used for Knowledge Sharing*. Stanford Knowledge Systems Laboratory (1993)
- [9] Guarino N., Giaretta P.: Ontologies and knowledge bases, towards a terminological clarification. *Towards very large knowledge bases: knowledge building and knowledge sharing*, IOS Presse, 25-32 (1995)
- [10] Knublauch, H. :Editing Semantic Web Content with Protégé the OWL Plugin. 6 the Protg workshop. Manchester, United Kingdom (2003)
- [11] Nabli,A., Feki, J., Gargouri, F.: An Ontology Based Method for Normalisation of Multidimensional Terminology, *Second International Conference on Signal-Image Technology and Internet-Based Systems*, (SITIS'2006), LNCS, 235-246 (2006)
- [12] Romero, O. and Abello, A.: Automating multidimensional design from ontologies. *Proceedings of the ACM international workshop on Data warehousing and OLAP*, 1-8 (2007)
- [13] Spyns, P., Meersman, R., Jarrar, M. : Data modelling versus ontology engineering. *SIGMOD Record*, 31(4), 12-17 ( 2002)
- [14] Stuckenschmidt, H., Parent, C., Spaccapietra, S. : Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization, *Lecture Notes in Computer Science* Vol. 5445 Springer 2009
- [15] Sugumaran, V., Storey, V.C., The role of domain ontologies in database design: An ontology management and conceptual modeling environment. *ACM Transactions on Database Systems*, 31(3), 10641094 (2006)
- [16] Yu, Z., Zhou, X., Hao, Y., Gu, J.: TV Program Recommendation for Multiple Viewers Based on user Profile Merging. *User Modeling and User-Adapted Interaction* 16(1), 63- 82 (2006).