COMBINING SEMANTIC TECHNOLOGIES AND DATA MINING TO ENDOW BSS/OSS SYSTEMS WITH INTELLIGENCE

Particularization to an International Telecom Company Tariff System

Javier Martínez Elicegui, Germán Toro del Valle
Telefónica Investigación y Desarrollo, Parque Tecnológico de Boecillo, Boecillo, Valladolid, Spain

Marta de Francisco Marcos
Alten Soluciones, Parque Tecnológico de Boecillo, Boecillo, Valladolid, Spain

Keywords: Semantic technologies, Data mining, BSS, OSS, Telecom, Tariff.

Abstract: Businesses need to "reduce costs" and improve their “time-to-market" to compete in a better position. Systems must contribute to these two goals through good designs and technologies that give them agility and flexibility towards change. Semantics and Data Mining are two key pillars to evolve the current legacy systems towards smarter systems that adapt to changes better. In this article we present some solutions to evolve the existing systems, where the end user has the possibility of modifying the functioning of the systems incorporating new business rules in a Knowledge Base.

1 INTRODUCTION

Companies in general and telecom operators in particular face a constant need to reduce costs and improve their time-to-market to compete in a better position in today’s highly competitive global market. Informational systems in the shape of OSS (Operational Support Systems) and BSS (Business Support Systems) contribute to these goals (Turban et al, 2006) providing the tools needed to increase the efficiency, agility and flexibility of organizations as a way to adapt themselves to a continuously changing environment.

In spite of the usefulness of these informational systems, the complexity of the current market landscape has caused an important decrease on their effectiveness requiring new techniques and technologies to recoup the lost space.

Semantic technologies and data mining constitute two key pillars to evolve the current legacy informational systems towards smarter ones which let us improve their adaptation to the current business environment.

In this paper we detail the procedure followed in Telefónica Investigación y Desarrollo (the R+D company of the Telefónica Group) as well as the results obtained from the application of these techniques and technologies to a critical informational system of the Telefónica Group as it is its tariff system.

The structure of the paper is as follows. The first section of the paper constitutes this introduction. In the second section, we introduce the tariff system used by Telefónica and the main problems it faces nowadays. The third section details the solution proposed with a special enfasis on the techniques and technologies applied as well as the results obtained. The forth section is devoted to the final conclusions.

2 THE TARIFF SYSTEM OF THE TELEFÓNICA GROUP

Gathering information from switchboards and billing telecom operator clients in real-time as they use the network, keeping updated on a second-by-second basis their consumption, is a complex challenge every telecom operator has to face. This complexity increases in the presence of highly heterogeneous

Martínez Elicegui J., Toro del Valle G. and de Francisco Marcos M. (2010).
COMBINING SEMANTIC TECHNOLOGIES AND DATA MINING TO ENDOW BSS/OSS SYSTEMS WITH INTELLIGENCE - Particularization to an International Telecom Company Tariff System.
In *Proceedings of the 12th International Conference on Enterprise Information Systems - Information Systems Analysis and Specification*, pages 350-355
DOI: 10.5220/0002975603500355

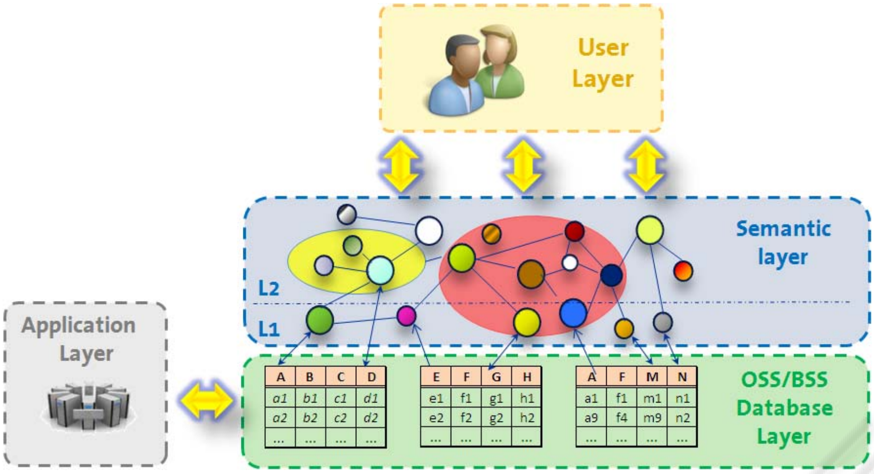

Figure 1: High level design of the proposed solution.

networks composed of equipments from distinct providers with a need to homogenize and to standardize the information provided by them, to relate this information to each phone call or data connection and to calculate the applicable cost.

Calculating the tariff to apply on second-by-second basis is probably one of the most complex parts of the whole billing process, since the tariff system makes it possible to define thousands of tariff schemas based on many criteria like, for example, the type of contract, promotions, origin and destination of the call, day of the week and time, duration of the call, etc.

To manage this complexity, Telefónica Investigación y Desarrollo has implemented a tariff system for mobile networks which is currently deployed in more than 15 countries from Europe and South America. Although this spread is a direct consequence of the success of this solution, it also becomes a major problem since the idiosyncrasy of each country raises particular needs which complicates the software versioning and maintenance.

To avoid the need to maintain a particular version of the tariff system for each country or area, the system has been designed as a highly parametrized system able to be configured and adapted to each particular case and installation, covering not only current needs but needs to appear in the future.

However, the level of granularity of this parametrization has its drawbacks since its management becomes a highly complex and specialized task as well as a not inconsiderable source of errors.

This situation combined with the increasing existent gap between business people (closer to the markets) and technical ones (closer to the particular implementation of the tariff system) made it necessary to rethink the problem and to face it using new techniques and paradigms which, on the one hand, made it easier to manage such a complex system and, on the other hand, brought the system closer to its final users (typicaly, business people in charge of defining the appropriate tariff schemas).

3 PROPOSED SOLUTION

To solve the aforementioned problem, the proposed solution consisted on building an Intelligent Configuration Assistant to exploit the inherent knowledge existent in the tariff system as a way to expose this knowledge directly to business people for their consumption and manipulation.

To build such a system, we leaned on semantic technologies (W3C: Semantic Web Case Studies) (Baader et al, 2004) (Davies et al, 2002), as a way to model this knowledge in a language close to business for its subsequent mapping to the more technical and parametrized language used by the final system.

The proposed solution, depicted in Figure 1, is based on four main layers:

- **OSS/BSS Database Layer.** corresponds to the tariff system configuration database.
- **Semantic Layer.** corresponds to a level on top of the tariff database. It includes a first level of concepts (L1) directly associated to the parameters stored in the database from which a second level of concepts (L2), more abstract and closer to business, can be defined

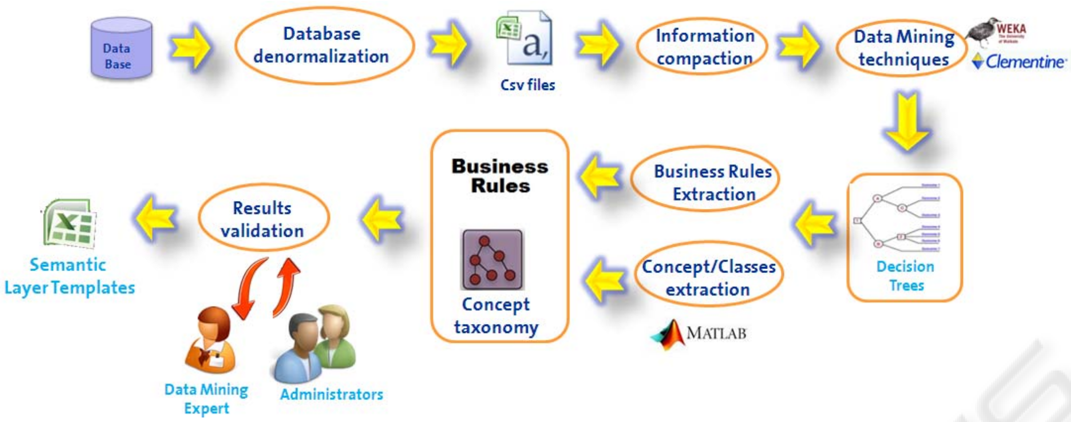

Figure 2: The knowledge extraction process algorithm.

by the administrators of the tariffs.

- **User Layer.** corresponds to the administrators of the tariff system who are able to define the business logic or rules amongst the concepts previously mentioned. The availability closer to business more abstract concepts makes it possible to define and to model business logic using not technical languages.

- **Application Layer.** corresponds to legacy applications which use the configuration parameters database during their execution to materialize the business concepts and rules defined. In this sense, a major requirement has been to provide a non-intrusive solution from a performance and current systems impact perspective.

From a practical viewpoint, the administrators are provided with Microsoft Excel templates which guide them through the definition of the Semantic Layer. Basically, they are able to define an ontology of high-level concepts based on already existent configuration parameters. Using a very simple syntax, administrators are also able to define the axioms which materialize the business rules which relates these high-level concepts to each other.

One of the most important difficulties of this process is providing the administrators with a preliminary version of the semantic layer based on the already existent configuration parameters.

Although this knowledge or model currently exists in the tariff system administrators minds as a way to translate the business schemas and requirements they receive to concrete configuration parameters stored in the system database, trying to extract and model this knowledge becomes an unapproachable task, mainly because of the reticences of the administrators themselves to expose the knowledge they have spent years collecting.

As a way to solve this *cold start dilema*, we have turned to data mining techniques as a way to extract a layer of semantics from the currently existent configuration parameters. These data mining techniques have allowed us to analyse the content of the databases used by the tariff system to extract a preliminary layer of inherent concepts, patterns and rules.

3.1 Knowledge Extraction through Data Mining Techniques

Identifying the inherent patterns, concepts and rules present in any database is a complex and laborious task consisting on a reverse engineering process applied to the design and contents of a concrete database.

At a first glance, this does not seem to be a hard problem to solve when we have a normalized database. Nevertheless, the problem increases when the database is populated with concrete data since from this population new subconcepts, patterns, rules and relationships (not even considered by the database designers) arise. For example, although the database may explicitly model the concept “Contract”, after populating the database with real data many other implicit concepts may arise like, for instance, “YoungContracts”, “ImmigrantContacts”, etc., which have to be modeled as well as their relationships to other implicit and explicit concepts exposed by the database. For example: “contracts by immigrants should be charged 0€ when the call is made to the country they were born in and 0,15€ to the rest”.

To solve this problem and to detect explicit and implicit concepts, patterns, rules and relationships

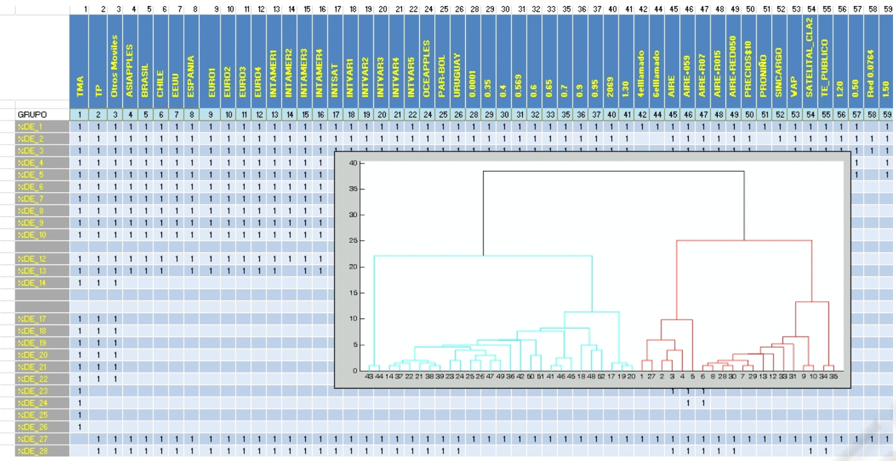

Figure 3: Concept taxonomy after applying clustering techniques.

exposed by the database, a 8-steps process has been proposed (see Figure 2):

1. **Database Denormalization.** the first step consists on generating the files where the data mining techniques will be applied. This step requires combining information spread amongst distinct tables of the database as a consequence of the database normalization design process. Sequences of queries are executed to combine fields of distinct tables and generate the desnormalization files.

2. **Information Compaction.** many entries of the database only differ in some easily identifiable concrete values. These similar registries can be grouped and treated as a set, reducing the data volume dramatically and increasing the efficiency of the data mining techniques to be applied afterwards. For example: entries which only differ in the time value can be grouped as one concrete instance assigned to a period of the day (i.e., “from 8:00 AM to 17:00 AM”).

3. **Data Mining Techniques.** once the information has been properly processed, it is analysed using distinct data mining techniques (Hair et al, 1995) (Fayyad et al, 1996). Initially, we used Weka data mining software (*http://www.cs.waikato.ac.nz/~ml/weka*) (Witten and Frank, 2005), more concretely the BFTree and J48 classification algorithms, but the high volume of data (more than 100.000 registries) and the complexity of these registries (more than 100 fields) made it unusable since the processing of the information went on for several days and frequently aborted because of unknown reasons. As a way to solve these performance issues, we moved to the SPSS’ Clementine solution (*http://www.spss.com*), obtaining much better results. In this sense, the C5 classification algorithm showed itself as a highly efficient algorithm to get optimum decision trees.

4. **Concept Extraction.** the decision trees obtained from the previous step expose data patterns. We can find sets of values, systematically repeated in different tree nodes, which denote some kind of semantic relation amongst them. From this assumption, we automatically process the decision trees searching for these data patterns and model them semantically in the shape of concepts (a.k.a. classes) which include the semantically related values (a.k.a. instances). The concepts defined in this step are manually validated by an administrator at a later extent, assigning them more precise business names.

5. **Concept Taxonomy.** after the previous step, we typically get a huge list of concepts which can hardly be assimilated all together by the administrators of the tariff system. To facilitate its understanding, we apply clustering and graphical representation techniques to the data (i.e. concepts) using tools such as *pdist* and *dendrogram* Matlab (*http://www.mathworks.com*) utilities (see Figure 3).

6. **Business Rules Extraction.** after the previous analysis to extract and to organize concepts, the decision trees are automatically processed

as business rules easier understandable by the administrator (i.e., Rule: IF contract= %TT23 AND destination=%DE62 AND day= (2) AND time = %TM12 …. THEN Q2=0*)*
The concepts extracted from the previous steps are used to model these business rules as a way to obtain more compact rules. These concept names are still cryptic (i.e. %DE62) and need to be renamed by business names.

7. **Result Validation.** at this point, the results obtained from all the automatic previous steps, in the shape of a taxonomy of concepts and a set of business rules, have to be validated jointly by the data mining experts and the administrators of the tariff system. This step makes it possible to detect errors as well as to model the knowledge extracted from the system database using more natural mechanisms (i.e. decomposing concepts into new subconcepts more familiar to the administrator of the tariff system, renaming concepts by business names, etc.), solving the cold start problem previously stated.

8. **Excel Templates Generation.** Finally, all the concept and business rule definitions are written down into the Excel templates to guide administrator during Semantic Layer maintenance.

3.2 Knowledge Modelling using Semantic Technologies

Modelling the inherent knowledge managed by the tariff system using semantic technologies (i.e., high-level concepts and business rules) highly simplifies the work of the administrators of the system and reduces the gap between them and the business experts, minimizing the probability of misunderstandings. The initial administrators job consisting on manually fine-tuning hundreds of parameters (in which case the business logic is implicit in the data managed) has been replaced with simple high-level business rules manipulations using Excel templates (in which case the data logic is explicitly exposed). This way, the administrators can easily adapt the tariff system to concrete business needs, considerably reducing not only the time-to-market but also the errors introduced into the system.

From this point, a whole new process start whose main aim is, on the first hand, to validate the specifications defined by the new semantic layer and, on the second hand, to act accordingly uppon the tariff system parameters database to expose this new behavior. This process is outlined in Figure 4 and consists on the following steps:

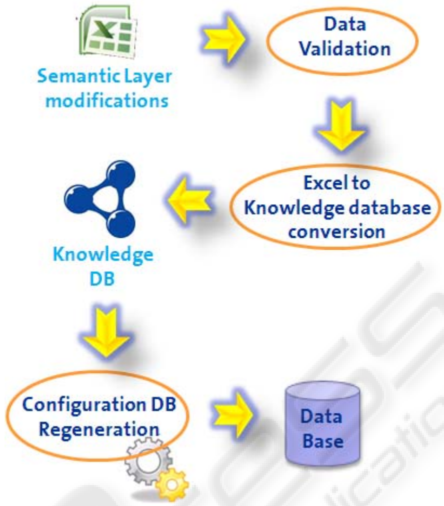

Figure 4: Knowledge modeling schema.

1. **Data Validation.** The new version of the semantic layer is validated verifying that the the business rules syntax is correct, and consistent with tariff system database.

2. **Excel to Knowledge Database Conversion.** once the semantic layer has been validated, its content is converted from human syntax (business rules) to machine syntax (OWL) (McGuiness and van Harmelen, 2004). This translation is automatically made using a set of UNIX scripts which fill with content previously prepaired OWL templates.

3. **Configuration Database Regeneration.** once a valid OWL ontology is available expressing the new concepts and business rules, a mapping of this ontology to the configuration detail parameters stored in the tariff system database has to be acomplished. For it, a set of queries allows us to extract the business logic expressed by the ontology and map it to concrete parameters of the database.

3.3 Lessons Learnt

One of the main problems we have faced during this project has been trying to improve the performance of the inference processes. In fact, we managed to reduce the initial 28 hours process to less than 1 hour.
To illustrate the problem and the solution proposed, consider the following business rule (see Figure 5),

where Q2 is a configuration parameter used by the tariff system.

Figure 5: Rule with premises as unique blank node express in Protégé (http://protege.stanford.edu/).

Using this kind of modeling, we obtained very bad performance during the inference phase, even if we reduced the data volume to 10% of the total.

A subsequent study of the reasons of this bad performance let us find out that each rule restriction was represented as a unique blank node (i.e, has_hours some Rush_hours). As a consequence, in case an instance is a member of one of these restrictions, it is also member of all blank nodes with the same logic in the rest of rules. This issue causes a larger number of anonymous clases and worst performance.

To avoid this behavior, we assign a unique named class to each restriction(i.e. RSV_Rush_Hours ≡ has_hours some Rush_Hours).

This modelling technique let us significantly improve the performance of the inference process (66% time reduction) as well as reduce the data volume to manage (50% reduction of number of triples).

A second aspect which let us significantly improve the performance of the whole process was to apply a “divide and conquer” strategy when feeding the assistant with data. We observed that feeding the assistant with a hundred of thousands instances significantly decreased the performance of the whole system. As a consequence, we divided the data into chunks of 15.000 instances and since the processing of each particular tariff instance is independent from the rest we processed the information in batch getting improvements in the processing time in a factor of 10.

4 CONCLUSIONS AND FUTURE WORK

In this article we have elaborated on some of the possibilities that semantic technologies and data mining offer to endow OSS/BSS systems with intelligence.

These technologies have a sufficient maturity level to be applied successfully to current legacy systems to provide a semantic layer on top of current databases.

As a particular case, we have applied this tecniques to the tariff system of the Telefónica Group where ongoing tariff calculations use the existing tables but a semantic layer on top of it helps us maintain the values of these tables up-to-date and consistent.

The main benefits of the implemented solution are:

- **Explicit Knowledge:** the tariff logic is now explicit, easily verifiable and editable by administrators.

- **Ease of Maintenance:** the knowledge managed by the system is now expressed in the shape of business rules. A simple change in one of these business rules may affect hundreds or thousands of records in the tariff tables with the certainty that the effect will be the desired one.

- **Risk Control:** expressing the knowledge managed by the system using formal semantic technologies allows us to automatically detect inconsistencies amongst rules, which prevent many of the current errors.

The experience and results obtained from this project encourages us to move forward and apply these same data mining and semantic techniques to other OSS/BSS systems of the company.

REFERENCES

Turban, E. et al, 2006. *Decision Support and Business Intelligence Systems*. Prentice-Hall, Inc.

Semantic Web Case Studies and Use Cases. W3C, (http://www.w3.org/2001/sw/sweo/public/UseCases)

Baader,F., Horrocks, I., Sattler, U., 2004. *Handbook on Ontologies*, Springer.

Davies, J., Fensel, D., van Harmelen, F., 2002. *Towards the Semantic Web: Ontology-driven Knowledge Management*. John Wiley and Sons, Inc.

Hair, J. F., Anderson, R. E., Tatham, R. L., and Black, W. C., 1995 *Multivariate Data Analysis (4th Ed.): with Readings*. Prentice-Hall, Inc.

Fayyad, U.,M. et al, 1996. *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press.

Witten, Ian H., Frank Eibe. 2005. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers.

McGuiness, D. L., van Harmelen, F., 2004. *OWL Web Ontology Language Overview*. W3C Recommendation. (*www.w3.org/TR/owl-features*)