Improving the knowledge discovery process using ontologies

Laurent Brisson, Martine Collard, Nicolas Pasquier

Improving the Knowledge Discovery Process Using Ontologies

Laurent Brisson, Martine Collard, Nicolas Pasquier
Laboratoire I3S
Université de Nice
2000 route des Lucioles
06903 Sophia-Antipolis, France
Email: {brisson,mcollard,pasquier}@i3s.unice.fr

***Abstract*— In this paper, we present the new ontology-based methodology ExCIS (Extraction using a Conceptual Information System) for integrating expert prior knowledge in a data mining process. This methodology describes guidelines for a data mining process like CRISP-DM. Its originality is to build a specific conceptual information system related to the application domain in order to improve datasets preparation and results interpretation. In this paper we specially present the CIS construction which consists of creating an ontology by information extraction from an initial raw database and building data to be mined.**

I. Introduction

One important challenge in data mining is to extract interesting knowledge and information useful for expert users. Numerous algorithms have been built for extracting best models according to quality criteria like accuracy, ROC area, lift and other indexes. Numerous quality measures have been proposed for objective and quantitative interest. Other works have focused on more semantic approaches for evaluating the subjective quality of discovered models. For instance, a dependency rule, association rule or classification rule, may be defined as interesting if it is either surprising or actionable [8], [13]. Thus the use of prior knowledge may significantly enhance the discovery of interesting patterns by considering the interestingness according to expert beliefs. In most data mining projects, prior knowledge is implicit only or it is not organized as a structured conceptual system. ExCIS is dedicated to data mining situations in which the expert knowledge is crucial for the interpretation of mined patterns, no conceptual representation of this knowledge is stored and only operational databases are given to the data mining process. In this approach, an ontology of the domain is built by analyzing existing databases with the collaboration of expert users who play a central role. The design process of the ontology is directed in order to facilitate the preparation of datasets and the interpretation of extracted patterns. The main objective in ExCIS is to propose a framework in which the extraction process makes use of a well-formed conceptual information system (CIS) for improving the quality of mined knowledge. We consider the paradigm of CIS as defined by [17]. The CIS provides the structure of information useful for further mining tasks. It contains: A conceptual schema defining an *Ontology* enhanced with a *Knowledge Base* (set of factual informations) on the specific domain of interest and a *Mining-Oriented relational DataBase* (MODB). Both ontology elicitation and database construction are mining-oriented; in this context, mining-oriented means that they are driven in order to facilitate knowledge discovery tasks.

Ontologies [4] are generally used to specify and communicate domain knowledge. They are formal, explicit specifications of shared conceptualizations of a given domain. Ontologies are very useful for structuring and defining the meaning of the metadata terms that are currently collected inside a domain community. They are a popular research topic in knowledge engineering, natural language processing, intelligent information integration and multi-agent systems. Ontologies are also applied in the World Wide Web community where they provide the conceptual underpinning for making the semantics of metadata machine understandable. While ontologies may be useful for conducting extraction in data mining tasks for discovering patterns, interpreting rules or conceptual clustering, they are not really integrated in current knowledge discovery projects.

In ExCIS, the ontology provides a conceptual representation of the application domain mainly elicited by analyzing the existing operational databases. ExCIS main characteristics are:

- Prior knowledge conceptualization: the CIS is specially designed for data mining tasks with the ontology, the prior expert knowledge base, the mining-oriented database.
- Adaptation of the CRISP-DM methodology with:
  - CIS based preparation of data sets to be mined.
  - CIS based post processing of mined knowledge in order to extract surprising and/or actionable knowledge.
  - Incremental evolution of the expert knowledge stored in the CIS.

In this paper, we present a kind of Customer Relationship Management (CRM) case study in which we apply the ExCIS methodology. This project deals with data from the 'family' branch of the French national health care system (CAF: Caisse Nationale d'Allocations Familiales). In this system, beneficiaries receive allowances depending on their social situation. The issue we address is to improve relationships between

Z.W. Ras, S. Tsumoto and D.A. Zighed (Eds) : IEEE MCD05, published by
Department of Mathematics and Computing Science Saint Marys University
Pages 25-32, ISBN 0-9738918-8-2, Nov. 2005

beneficiaries and the CAF organism. In this case study, we had two sources of information: an operational database storing data on beneficiaries and their contacts with the CAF was provided prior knowledge of expert users aware of the business processes, behaviors and habits in the organism.

The paper is organized as follows. Section 2 gives an overview of the ExCIS approach. In Section 3, we study related works. Section 4 describes the underlying conceptual structures of the ontology. In Section 5, we give a detailed description of CIS construction. Section 6 concludes the paper.

II. Overview of the ExCIS Approach

ExCIS integrates prior knowledge all along the mining process: the first step structures and organizes the knowledge in the CIS and further steps exploit it and enrich it too.

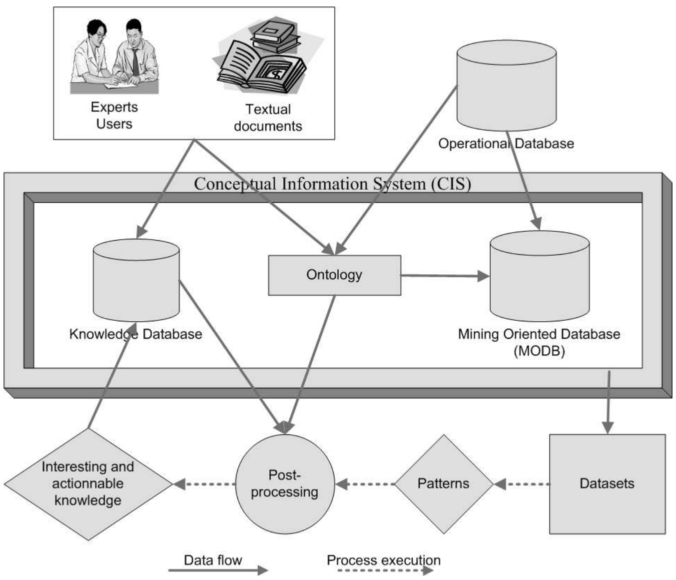

Fig. 1. Hierarchical clustering of conditions.

The global ExCIS process presented in 1 shows:

- The CIS construction where:
  - The ontology is extracted by analyzing operational databases and by interacting with expert users.
  - The knowledge base, set of factual informations, is obtained in a first step from dialogs with expert users.
  - The new generic MODB is built.
- The pre-processing step where specific datasets may be built for specific mining tasks.
- The standard mining step which extracts patterns from these datasets.
- The post-processing step where discovered patterns may be interpreted and/or filtered according to both prior knowledge stored in the CIS and individual user attempt.

The MODB is said to be generic since it will be use as a kind of basic data repository from which any task-specific dataset may be generated. The underlying idea in the CIS is to build structures which will provide more flexibility not only for pre-processing the data to be mined, but for filtering and interpreting discovered patterns in a post-processing step. Hierarchical structures and generalization/specialization links between ontological concepts play a central role:

- They allow reducing the volume of extracted patterns like sets of rules which are often very large.
- They provide a tool for interpreting results obtained by clustering algorithms too.

For numerical or categorical data, they provide different granularity levels which are useful in the pre-processing and the post-processing steps.

a) Example 1: If we assume that 10, 11, 12 and 13 are defined in the ontology as concepts (day numbers) which inherits (are more specialized) from the more general concept WeekNumber = 2, then Rule1, Rule2, Rule3, Rule4 may be replaced by the only rule Rule5 which left hand side is more general.
Rule1: If DAYNUMBER = 10 and REASON = ‘Entering Call’ then OBJECTIVE = ‘To Be Paid’
Rule2: If DAYNUMBER = 11 and REASON = ‘Entering Call’ then OBJECTIVE = ‘To Be Paid’
Rule3: If DAYNUMBER = 12 and REASON = ‘Entering Call’ then OBJECTIVE = ‘To Be Paid’
Rule4: If DAYNUMBER = 13 and REASON = ‘Entering Call’ then OBJECTIVE = ‘To Be Paid’
Rule5: If WEEKNUMBER = 2 and REASON = ‘Entering Call’ then OBJECTIVE = ‘To Be Paid’

b) Example2: Let us consider the following example on categorical data: if ‘Student Housing Allowance’ and ‘Family Housing Allowance’ are defined in the ontology as instances of concept ‘Housing Allowances’ and if the two following association rules Rule6 and Rule7 are extracted, a generalization may be applied to replace these two rules by the more general rule Rule8.
Rule 6: If PLACE = ‘Main Office’ and REASON = ‘Documents’ and RESULT = ‘Intervention Dossier’ then OBJECTIVE = ‘Student Housing Allowance’
Rule 7: If PLACE = ‘Main Office’ and REASON = ‘Documents’ and RESULT = ‘Intervention Dossier’ then OBJECTIVE = ‘Family Housing Allowance’
Rule 8: If PLACE = ‘Main Office’ and REASON = ‘Documents’ and RESULT = ‘Intervention Dossier’ then OBJECTIVE = ‘Housing Allowances’
Techniques for integrating prior knowledge as illustrated by previous examples may be introduced at the post-processing steps of the data mining process.

III. Related Works

A. Interestingness Measures

Numerous works focused on indexes that measure the interestingness of a mined pattern [5], [9]. They generally distinguished objective and subjective interest. Among these indexes there are quantitative measures of objective interestingness such as confidence, coverage, lift, success rate while unexpectedness and actionability are proposed for subjective criteria. Since our work deals with user interestingness, we focus this state of the art on the former. According to the actionability criteria, a model is interesting if the user can start some action depending on it [14]. On the other hand, unexpected models are considered interesting since they contradict user expectations which depend on his beliefs.

User expectations is a method developed by Liu [9]. The first approach neither dealt with unexpectedness nor actionability. User had to specify a set of patterns according to his previous knowledge and intuitive feelings. Patterns had to be expressed in the same way that mined patterns. Then Liu defined a fuzzy algorithm which matches these patterns. In order to find actionable patterns, the user has to specify all actions that he could take. Then, for each action he specifies the situation under which he is likely to run the action. Finally, the system matches each discovered pattern against the patterns specified by the user using a fuzzy matching technique.

Silberschatz and Tuzhilin [13] proposed a method to define unexpectedness via belief systems. In this approach, there are two kinds of beliefs: soft beliefs that the user is willing to change if new patterns are discovered and hard beliefs which are constraints that cannot be changed with new discovered knowledge. Consequently this approach assumes that we can believe in certain statements only partially and some degree or confidence factor is assigned to each belief. A pattern is said to be interesting relatively to some belief system if it ‘affects’ this system, and the more it ‘affects’ it, the more interesting it is. However, interestingness of a pattern depends also on the kind of belief.

Piatetsky-Shapiro [11] presented KEFIR, a discovery system for data analysis and report generation from relational databases. This system embodies a generic approach based on the discovery technique of deviation detection [10] for uncovering ‘key findings’, and dependency networks for explaining the causes of theses findings. Central to KEFIR’s methodology is its abilities to rank deviations according to some measure of interestingness. Interestingness in KEFIR refers to the degree to which a discovered pattern is of interest to the user of the system and is driven by factors such novelty, utility, relevance and statistical significance [3].

B. Databases and Ontologies

Ontologies provide a formal support to express beliefs and prior knowledge on a domain. Domain ontologies are not always available; they have to be built specially by querying expert users or by analyzing existing data. Extracting ontological structures from data is very similar to the process of retrieving a conceptual schema from legacy databases. Different methods [7], [6], [16], [12] were proposed. They are based on the assumption that sufficient knowledge is stored in databases for producing an intelligent guide for ontology construction. They generally apply a matching between ontological concepts and relational tables such that the ontology extracted is very close to the conceptual database schema.

C. Ontologies and Data Mining

For the last ten years, ontologies have been extensively used for knowledge representation and analysis mainly in two domains: Bioinformatics and web content management. Biological knowledge is nowadays most often represented in ‘bio-ontologies’ that are formal representations of knowledge areas in which the essential terms are combined with structuring rules that describe relationships between the terms. Bio-ontologies are constructed according to textual descriptions of biological activities. One of the most popular bio-ontology is *Gene Ontology*[1] that contains more than 18 thousands terms. It describes the molecular function of a gene product, the biological process in which the gene product participates, and the cellular component where the gene product can be found. Results of data mining processes can then be linked to structured knowledge within bio-ontologies in order to explicit discovered knowledge, for instance to identify biological functions of genes within a cluster. Interesting surveys of ontologies usage for bio-informatics can be found in [1], [15]. A successful project of data mining application using bio-ontologies is described in [18].

In the domain of web content management, OWL (Ontology Web Language)[2] is a Semantic Web standard that provides a framework for the management, the integration, the sharing and the reuse of data on the Web. Semantic Web aims at the sharing and processing of web data by automated tools as well as by people. It can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms, i.e. an ontology. Web ontologies can be used to enrich and explain extracted patterns in many knowledge discovery applications to web such as web usage profiling [2] for instance.

IV. Conceptual structures of the ontology

A. Ontology

In EXCIS the domain ontology is an essential means both for improving data mining processes and for interpreting data mining results. The ontology is defined by a set of concepts and relationships among them which are discovered by analyzing existing data. It provides support both in the pre-processing step for building the MODB and in the post-processing steps for refining mined results.

As shown in section II, generalization/specialization relationships between ontological concepts provide valuable information since they may be used intensively for reducing and interpreting results. For instance, a set of dependency rules (attribute-value rules) may be reduced by generalization on attributes (see example 1 above) or by generalization on values (see example 2 above). Thus the guidelines in the ontology construction are:

- To distinguish attribute-concept and value-concept.

[1] http://www.geneontology.org/
[2] http://www.w3.org/TR/owl-ref/

- To establish matching between source attributes and attribute-concepts and a matching between source values and value-concepts.
- To define concept hierarchies between concepts.

This ontology does not contain any instances since values are organized in hierarchies and considered as concepts. In EXCIS, according to the data mining orientation, the ontology has two important characteristics:

- The ontology does not contain any instance since source values are organized in hierarchies and considered as concepts; instances are only present in the final relational database MODB.
- Each concept has two generic properties only.

The MODB is a relational database whose role is to store the most fine-grain data elicited from the operational database. MODB attributes are those which are identified as relevant for the data mining task and MODB instances are tuples of most fine-grain values.

B. Concept definition

In a EXCIS ontology, a concept has to refer to a domain paradigm useful in the data-mining process. A concept in EXCIS is characterized by the two following properties: Its role in an extracted pattern (attribute/value) and a boolean property (abstract/concrete) which indicates its presence in the final MODB. An attribute-concept is a data property, and a value-concept represents values of a data property. For instance in figure 2 "Children number" is an abstract attribute-concept and "3 children" is a concrete value-concept.

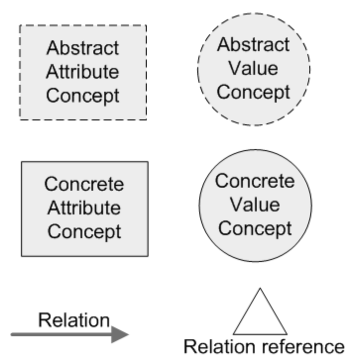

Fig. 2. Legend

A concept (attribute-concept or value-concept) which is in the MODB is called a *concrete concept*. A concept that is not concrete but is useful during the post-processing step is called an *abstract concept*.

C. Relationship definition

A relationship is an oriented link between 2 concepts. In ExCIS there are 4 different kinds of concepts and we distinguish relationships between concepts of the same hierarchy and concepts of different hierarchies: Thus there are 32 different kind of relationships between concepts. Among them we can set up 3 different categories:

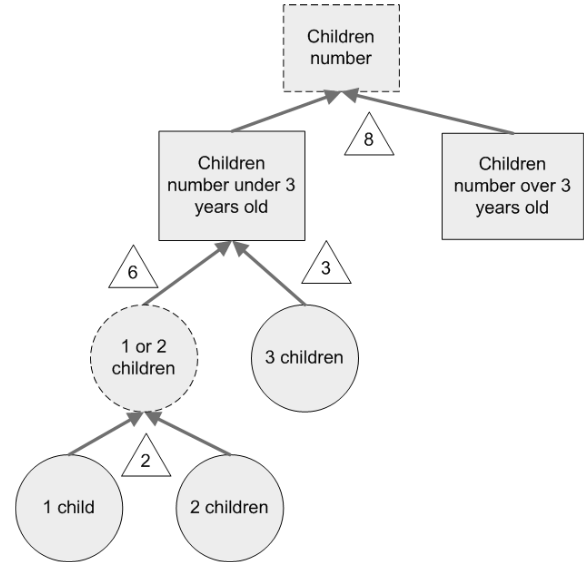

Fig. 3. Children related concepts

- Generalization/specialization relationships between two value-concepts which are "is-a-kind-of" links between two value-concepts (see relationship 2 in figure 3)
- Generalization/specialization relationships between two attribute-concepts which are "is-a-kind-of" links between two attribute-concepts (see relationship 8 in figure 3)
- Generalization/specialization relationships between a value-concept and an attribute-concept which are "is-a-value-of" links (see relationship 2 in figure 3)

All relationships between concepts within the same hierarchie are in table 1, and relationships between concepts in different hierarchies are in table 2. Numbers are relation references and forbidden relationships are indicated by letters.

TABLE I
CONCEPT RELATIONSHIPS WITHIN THE SAME HIERARCHIE

Concept	Concrete Value	Abstract Value	Concrete Attribute	Abstract Attribute
Concrete Value	1	2	3	D
Abstract Value	B	2	6	5
Concrete Attribute	C	C	9	7,8
Abstract Attribute	C	C	A	10

TABLE II
CONCEPT RELATIONSHIPS IN DIFFERENT HIERARCHIES

Concept	Concrete Value	Abstract Value	Concrete Attribute	Abstract Attribute
Concrete Value	4	4	D	5
Abstract Value	B	4	D	5
Concrete Attribute	C	C	D	D
Abstract Attribute	C	C	A	D

D. Description and use of ontology relationships

First of all, two relationships are forbidden into ExCIS: generalization/specialization relationships from an abstract concept toward a concrete concept with the same role (attribute or value) since we want abstract concepts to be more general than concrete concepts (see relationship A or B in figure 4), and generalization/specialization relationships from an attribute-concept toward a value-concept (see relationship C in figure 4) since the relationship "is a value of" has no meaning in this situation.

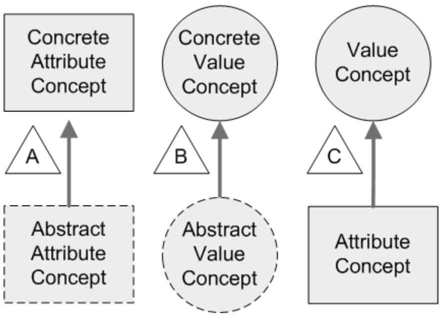

Fig. 4. Forbidden relationships

1) Relationships between value-concepts: Generalization or specialization relationships between value-concepts (see relationship 2 in figure 3) are useful in order to generalize patterns during the post-processing step. Furthermore, relationships between two concrete value-concepts of the same hierarchy are essential since they also allow selecting data granularity in datasets generated from the MODB. If in a data mining session we are more interested in the kind of allowances than in specifics allowances, dataset granularity will be set to the "Housing Location" level (see relationship 1 in figure 5).

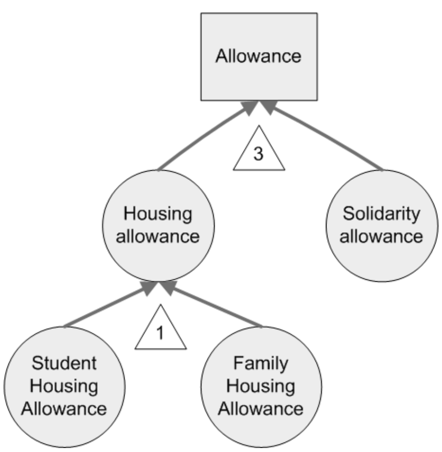

Fig. 5. Allowances related concepts

2) Relationships between attribute-concepts: Generalization or specialization relationships between attribute-concepts are useful in order to generalize models during the post-processing step. However, we must be careful during this generalization since sometimes we can switch an attribute with a more general (see relationship 7 in figure 6) and other

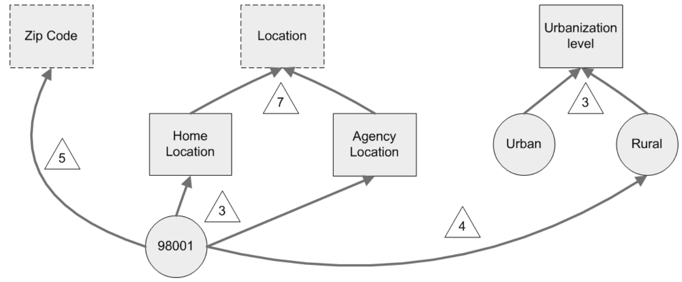

Fig. 6. Location related concepts

times we have to compute a new value (see relationship 8 in figure 3). For instance in figure 3 "Children number" is the sum of all of the values of his subconcepts.

Relationships between two concrete attribute-concepts of the same hierarchy are specific because they have to be checked during datasets generation: indeed these attributes cannot be in the same dataset to avoid redundancy.

ExCIS method forbids relationships between attribute-concepts of different hierarchies. Relationships between attribute-concepts are only generalization relations; attribute-concepts which are semantically close have to be located together in the same hierarchy.For instance "Home Location" and "Agency Location" concepts are in the "Location" hierarchie (see relationship 7 in figure 6).

3) Relationships between value-concepts and attribute-concepts: These relationships are essential in order to build data or to provide different semantic views during the post-processing step (see relationship 5 in figurein figure 6). For instance "98001" is both a "Home Location" and a "Zip Code" (see relationship 3 in figure 6).

Each value-concept is linked with attribute-concepts in the same hierarchy. ExCIS forbids relationships between a value-concept and concrete attribute-concepts of different hierarchies: indeed if such a relationship exists, it means that a value-concept "is a value of" two different attribute-concepts. If these attribute concepts are semantically close they must be in the same hierarchy and if they are totally different they can't be in relationship with the same value-concepts.

V. Conceptual Information System Construction

Let A the set of source database attributes, C the set of ontology concepts and C_z the set of concepts associated to an attribute $z \in A$. C is defined by $\bigcup_{z \in A} C_z$.

ExCIS differs from CRISP-DM mainly in the data preparation step. In this step CRISP-DM describe 5 tasks: select, clean, construct, integrate and format data. Selection and format are identical in both methods but in ExCIS cleaning, construction and integration are merged method in order to elicitate the ontological concepts and to build the MODB.

A. Scope Definition and Source Attribute Selection

First steps of ExCIS method are related to the Business Understanding and the Data Understanding steps of CRISP

DM method. They need an important interaction with expert users.

1. Determine objectives: in our case study, objectives are to improve "relationships with beneficiaries".
2. Define themes: analysis of data allow to gather them into semantic sets called themes. For example we create 3 themes: Allowance beneficiaries profiles, contacts (by phone, by mail, in the agency, ...) and events (holidays, school starts, birth, wedding, ...).
3. For each theme select a set of source attributes with experts users.

B. Data Analysis and Attribute-Concept Elicitation

4. For each selected attribute z:
   5. Examine name and description in order to:
   - Associate n concepts to the attribute.
   - Into C, clean homonyms (different concepts with same name), synonyms (same concepts with different names like age and date of birth) and useless attributes according the objectives.
   6. Examine values (distribution, missing values, duplicates values, ...) in order to:
   - Refine C_z (add or delete concepts) according to information collected in step 6.
   - Clean again homonyms, synonyms and useless attributes. For example by analyzing values we realized that 'allowances' was in fact 2 homonyms concepts. Thus we created the 'allowance amount' concept and the 'allowance beneficiary' concept.
7. For each concept associated to z create the method which generates value-concepts.

In the step 7, if the attribute-concept doesn't exist we have to create a 4 fields record table. These fields are the attribute associated to the concept, the name of the attribute table in source database, the attribute domain value and the reference to the procedure which may generate value-concepts. There is only one procedure for record in the table. A domain value can be a distinct value or a regular expression and is the input of the procedure. Procedure output provides references to value-concepts. The procedure might be an SQL request (SELECT or specific computation) or an external program (script, shell, C, ...). However, if the attribute-concept already exists we just have to add a record in the table and create a new procedure.

C. Value-Concept Elicitation

At this point, all of the methods to generate value-concepts are created.

8. Give a name to each value-concept.
9. Clean homonyms and synonyms among value-concepts.

D. Ontology Structuration

10. Identify generalization relationships among value-concepts (see figure 5).
11. Create abstract concepts and reorganize the ontology with these new concepts. For instance 'Location' concept in figure 6.
12. Create relationships between value-concepts of different hierarchies (see relation 4 in figure 6).

E. Generation of the Mining Oriented Database

13. Generate the database by using procedures defined in step 7.

In this final step a program reads the tables created for each attribute-concept and calls the procedures in order to generate the MODB.

VI. Conclusion

We have given a global presentation of the new methodology ExCIS for the integration of prior knowledge in a data mining process. The main objective is to improve the quality of extracted knowledge and facilitate its interpretation. ExCIS is based on a conceptual information system (CIS) which stores the expert knowledge. The CIS plays a central role in the methodology since it is used for datasets construction before mining, for filtering and interpreting mined patterns and for updating expert knowledge with validated mined knowledge. This paper focuses on the CIS structure and on its construction only. We have presented its ontological structures, and we have discussed the choices made for identifying ontology concepts and relations by analyzing existing operational data. Further works will be dedicated to the pre- and post-processing steps. We will study techniques for efficiently exploiting information from the domain ontology in order to conduct the preparation of datasets to be mined and the interpretation of mined results.

VII. Acknowledgements

We would like to thank the CAF (family branch of the French national health care system) and more specially Pierre Bourgeot, Cyril Broilliard, Jacques Faveeuw, Hugues Saniel and the BGPEO for supporting this work.

References

[1] J.B. Bard and S.Y. Rhee. *Ontologies in Biology: Design, Applications and Future Challenges*. Nature Review Genetics, 5(3):213-222, march 2004.
[2] H. Dai and B. Mobasher. *Using Ontologies to Discover Domain-level Web Usage Profiles*. Proceedings 2nd ECML/PKDD Semantic Web Mining workshop, august 2002.
[3] W.J. Frawley, G. Piatetsky-Shapiro and C.J. Matheus. *Knowledge Discovery in Databases: An Overwiew*. In Knowledge Discovery in Databases, pp. 1-27, AAAI/MIT Press, 1991. Reprinted in AI Magazine, 13(3), 1992.
[4] T. Gruber. *What is an Ontology?*, january 2002. http://www-ksl.stanford.edu/kst/what-is-an-ontology.htm.
[5] R.J. Hilderman and H.J. Hamilton. *Evaluation of Interestingness Measures for Ranking Discovered Knowledge*. Proceedings 5th PAKDD conference, Lecture Notes in Computer Science 2035:247-259, april 2001.
[6] P. Johannesson *A Method for Transforming Relational Schemas into Conceptual Schemas*. Proceedings 10th ICDE conference, M. Rusinkiewicz editor, pp. 115-122, IEEE Press, febuary 1994.
[7] V. Kashyap. *Design and Creation of Ontologies for Environmental Information Retrieval*. Proceedings 12th workshop on Knowledge Acquisition, Modelling and Management, october 1999.
[8] B. Liu, W. Hsu and S. Chen. *Using General Impressions to Analyze Discovered Classification Rules*. Proceedings 3rd KDD conference, pp. 31-36, august 1997.

[9] B. Liu, W. Hsu, L.-F. Mun and H.-Y. Lee. *Finding Interesting Patterns using User Expectations*. Knowledge and Data Engineering, 11(6):817-832, 1999.

[10] C.J. Matheus, P.K. Chan and G. Piatetsky-Shapiro. *Systems for Knowledge Discovery in Databases*. IEEE Transactions on Knowledge and Data Enginneering, 5(6):903-913, december 1993.

[11] G. Piatetsky-Shapiro and C. Matheus. *The Interestingness of Deviations*. Proceedings of the AAAI-94 workshop on Knowledge Discovery in Databases, 1994.

[12] D.L. Rubin, M. Hewett, D.E. Oliver, T.E. Klein, R.B. Altman *Automatic Data Acquisition into Ontologies from Pharmacogenetics Relational Data Sources using Declarative Object Definitions and XML*. Proceedings 7th Pacific Symposium on Biocomputing, pp. 88-99, january 2002.

[13] A. Silberschatz and A. Tuzhilin. *On Subjective Measures of Interestingness in Knowledge Discovery*. Proceedings 1st KDD conference, pp. 275-281, august 1995.

[14] A. Silberschatz and A. Tuzhilin. *What Makes Patterns Interesting in Knowledge Discovery Systems*. IEEE Transaction On Knowledge And Data Engineering, 8(6):970-974, december 1996.

[15] R. Stevens, C.A. Goble and S. Bechhofer. *Ontology-based Knowledge Representation for Bioinformatics*. Brief Bioinformatics, 1(4):398-414, november 2000.

[16] L. Stojanovic, N. Stojanovic and R. Volz. *Migrating Data-intensive Web Sites into the Semantic Web*. Proceedings 17th ACM Symposium on Applied Computing, pp. 1100-1107, ACM Press, 2002.

[17] G. Stumme. *Conceptual On-Line Analytical Processing*. K. Tanaka, S. Ghandeharizadeh and Y. Kambayashi editors. Information Organization and Databases, chpt. 14, Kluwer Academic Publishers, pp 191-203, 2000.

[18] N. Tiffin, J.F. Kelso, A.R. Powell, H. Pan, V.B. Bajic and W.A. Hide. *Integration of Text- and Data-Mining using Ontologies Successfully Selects Disease Gene Candidates*. Nucleic Acids Research, 33(5):1544-1552, march 2005.