International Journal of Computer Science and Applications,
2008, Vol. 5, No. 3b, pp 93 - 116

THE USE OF A FORMAL REPRESENTATION OF ACCOUNTING STANDARDS

PIERRE TELLER

I3S, University of Nice
2000 route des lucioles
06903 Sophia-Antipolis cedex, France

pierre.teller@i3s.unice.fr
http://<www.i3s.unice.fr/ pteller>

The purpose of this paper is to introduce the model we built to represent financial information in an efficient way. This kind of information shows an unquestionable delay on the field of knowledge management, and this issue is more and more crucial as the requirements on financial reporting increase. We will describe the particular nature of laws as a knowledge source, and amongst other, the nature of accounting standards. We will explain the consequences of these particularities on the technical choices we made when we started to build our model, and what is the structure of our model. Then, we will sum up the issues we faced during the modelling process. This paper ends with a description of the applications we build to illustrate the possibilities of such a model for various financial information users, especially financial auditors.

Keywords : description logics, ontologies, accounting standards, knowledge representation, XBRL

1. Introduction

The environment in which financial information is produced and analyzed is changing widely. All over the world, we can see a general trend which consists in more frequent, more extensive and more documented reporting. These changes are the result of several phenomenons, mainly the constantly growing volume of data to manipulate and the recent financial scandals (Enron, Worldcom, etc.) [Veron (2004)]. A set of international standards called IFRS (International Financial Reporting Standards) emerged during the last years. They are already used by public companies in the European Union, and their use should become widespread in the future.

Thus, we can see that the reporting requirements have considerably increased, both on a qualitative and on a quantitative way. To face this evolution, the use of computers becomes more extensive in data processing. These reasons gave rise to a new language : XBRL (eXtensible Business Reporting Language). This language consists in XML taxonomies, where each taxonomy is dedicated to a particular subset of business reports [Richards (2002a)] [Richards (2002b)](like a taxonomy for metallurgy industry, for tax reports, etc.) and aims to become a standard worldwide. The extensive use of this language would be an unquestionable step forward [Zarowin (2003)]. Moreover, the IFRS recommend companies to report using this language [ICCA (2002)].

Let us study more accurately what XBRL could bring to financial reporting.

1. **Production and publication of the information**: the use of a common language allows a strong reduction of the time spent to compile data from varied sources, by avoiding most of the reprocessing. It also reduces the risk of typing error, which would otherwise be multiplied with each reprocessing.
2. **Information sharing and comparability**: a common language makes it easier to understand pieces of information by external users. The mark-up language, if the tags are relevant, can describe the context in which the information was extracted and produced. This context may not be accessible with the raw information, for instance, a list of numbers representing money can not be exploited if the currency is not specified. That is the interesting point with XBRL, which inherits from XML the ability to tag information and make the information more comparable.
3. **Information checking and certification**: the tagging enables easier checking because the reader knows where each particular piece of information should be. It does not need to be searched or reconstructed, allowing the possibility to mistake pieces of information. Additionally the easier an information can be checked, the more it can be trusted, it is a guarantee of transparency
4. **Gain in the analysis process**: it is common knowledge that financial analysts spend more than a third of their time to reprocess the information, and therefore an important part of their work can not be dedicated to real analysis. Information that can be immediately utilized allows important time savings, a reduction of typing mistakes, and the possibility to analyze more information of better quality.

These issues are neither new nor characteristic of financial information, but would allow many cost-savings if widely applied [Microsoft (2005)]. But in our opinion, even if the apparition of this standard is important, it is not enough. XBRL is just an optimal way to store information, but there is an important work to do upstream (see figure 1). XBRL does not provide an explicit formalism of financial data, and more particularly of accounting standards. In fact, an implicit formalism must have been used when specifying XBRL taxonomies, but since it is not clearly published, it is not possible to use it for knowledge extraction and manipulation.

In other words, accounting professionals need modern and efficient information systems [ISACA (2002)], and these information systems require machine-readable data that allows reasoning, not only an electronic storage format. That is why we decided to build a formal representation of accounting standards. This is a major issue for financial information consumers. More and more companies do not have information systems that are able to face the new legal requirements (see [Henry (2007)] and figure 2).

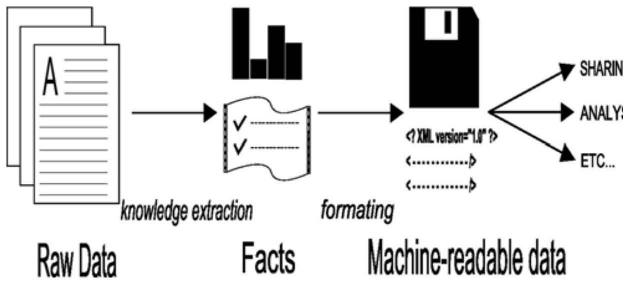

Figure 1. XBRL is only useful at the end of information chain

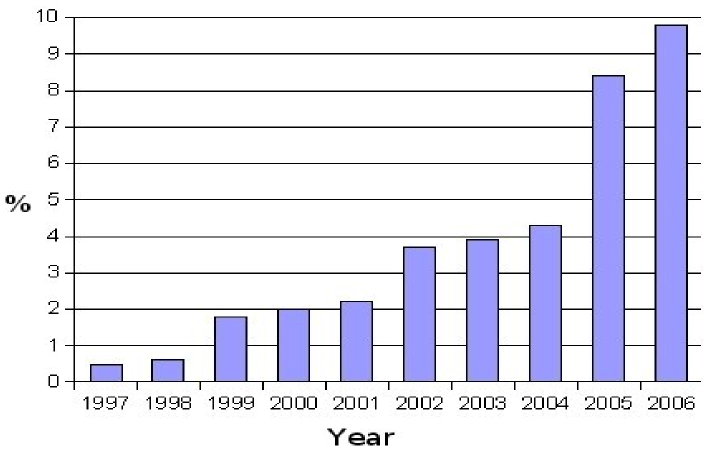

Figure 2. Percentage of financial statements refused by the administration in the US over the last ten years

In this paper, we will first describe the nature of financial information, and the main features of the accounting standards, which are a particular case of laws and regulations. The modeling of this kind of domain has particularities that are important to consider. Then, in the third chapter, we will explain the choices we made. In the following two chapters, we will describe our methodology and the results we have reached : a model of accounting standards composed of two different and complementary parts, a syntactic one and a semantic one. And finally, we will discuss the possible extensions of our works. We will describe how we have used our model in practical cases so far, and other possible extensions.

2. The Particular Nature of Financial Information

In order to build a suitable formalism for our model of representation, we must consider the nature of financial information, to see which are the particular constraints of this domain.

The legal context which describes financial information is rather difficult to represent. The definitions of the accounting rules are based on a conceptual framework, which only defines general principles: the information must be comparable, relevant, and give a faithful image of the entity that one wants to describe [Boussard (1997)]. A whole set of practices and habits, which are in a certain way implicit, lie beyond these general principles [Gensse (2000)]. As a consequence, it is often made reference to the ”interpretations” which have been made of the accounting rules. These interpretations are a kind of jurisprudence, their goal is to specify the meaning of an accounting definition in cases where the initial definition was ambiguous. These interpretations may change with time, and may be different from one country to another [Alia (2000)].

Above this conceptual framework, there are reference texts defining the accounting standards in each country or area (US GAAP in the United States, IFRS partly adopted in European Union, etc.) claiming to define precise rules. The validity of financial information is assessed through comparison with these texts. There should not be any contradiction between the conceptual framework and the reference texts.

However, because of the multitude of particular cases, it is not possible to exclude this possibility. Moreover, some IFRS rules admit the possibility of contradictions, and in this case they specify that the general principles must prevail.

This situation is common when trying to represent knowledge from laws and regulations [Kralingen (1995)]. The sources are constituted by various forms of documents, with different importance, and that’s why conflict management is a crucial issue in this domain. The representation must respect the original consistency of the sources, but it must also be able to deal with conflicts that could arise. Another important point is that the representation must be able to face potential changes: this domain is not static, and a modification of one document may impact many other documents that depend on it in the hierarchy. Finally, this is typically a domain where there is a great freedom of interpretation, and it is not always possible to decide if a particular interpretation will always be true. So the model must be able to take this uncertainty into account.

These specifications will impact our choices of methods and languages. For a study of the representation of a particular case of regulation context representation, see [Jouve (2003)].

3. Technical Choices

Our goal is to describe two complementary aspects of the accounting standards, the first one is syntactic, and the other one is semantic. Our model would be the sum of the two descriptions. The syntactic description emphasizes the global framework of the standards, how they are articulated, what is their hierarchy, what objects they impact and how they impact them. It will also decide what kind of inference it will be possible to make. This will be discussed in chapter 4. Semantic description will list all the elements affected by accounting standards, and the meaning of these elements will become explicit. This will be explained in chapter 5. Figure 3 represents the different steps.

Most of the studies on accounting are limited to this last part, i.e. purely semantic [Degos (2005)]. They do not emphasize a general structure and they behave like a dictionary, describing each element as a particular case. That is why the syntactic study is important to reach our goals. Conversely, a purely syntactic description is insufficient. Indeed, because of the nature of financial information, we must be able to give the exact meaning of every term used to describe the syntax, the context in which it takes place, the standards and the interpretations which refer to it, etc. We feel that it is essential that both descriptions be implemented. Moreover, we think that it is much better to build them simultaneously. First of all, this ensures coherence between the concepts used in the syntax and their meaning. Secondly, relations and dependencies between concepts must be taken into account in our syntax. So we decided to take a holistic approach, each task validating the other.

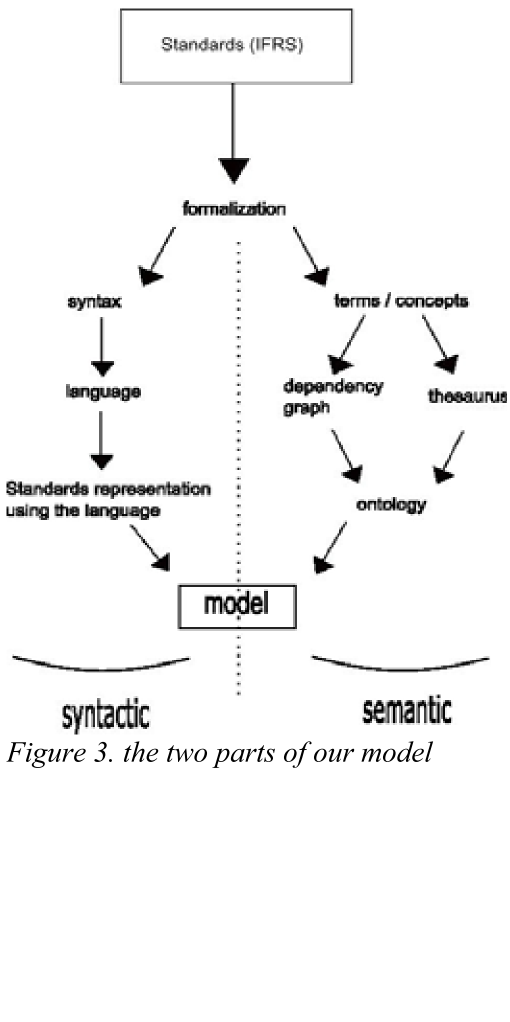

Figure 3. the two parts of our model

3.1. *For the Syntactic Part*

Amongst all the families of languages we have examined, description logics appeared to be the better for our purpose. First, they have high expressive power, as well as decidable and efficient inference mechanisms [Nardi (2003)]. Secondly, they have the interesting property that concepts can be defined by describing properties that an object must satisfy in order to belong to the concept. This is an important point, because many objects in the IFRS texts are defined like this. That is the reason why our syntactic model has common points with classic description logics. However, it also has some differences.

Nevertheless, we could not use directly the most classical forms of description logics, because some particular constraints of this domain implied some differences: the IFRS define some rules that are not decidable and that involve reader's judgmental, or even probabilistic analysis of context. This kind of definitions can not be described by a subset of first order logic, and thus our model is not a standard description logic. For this purpose, we used some logical operators inspired from deontic logic [Garson (2007)]. These operators were very adapted to this situation. It is very easy to introduce these operators with the description logic formalism, and that were a very important factor when we had to choose a formalism.

3.2. *For the Semantic Part*

The choice of a formalism for this part of the model was easier. We built an ontology, because this structure is very powerful for representing the knowledge of a domain with concepts and relations between these concepts. The relations lead to a subsumption hierarchy of the concepts which is by itself a representation of the domain. The description of the concepts allows to set the meaning attached to these concepts, and thus avoids misunderstanding with other users of our model.

The definition of what is exactly an ontology is still a source of debate. Nevertheless, some definitions are more and more popular, and a consensus appears slowly: an ontology is a representation of a shared conceptualization of a domain. There is a good description of ontologies and their predecessors in [Gandon (2006)]. The origin of this formalism is represented in figure 4.

Since this formalism is more and more popular, there are various tools to edit ontologies now. For a complete summary, see [Cardoso (2007)]. We used the editor Protégé, from Stanford University [Noy (2001)]. It is a powerful tool, and it lets a great freedom to the modeler. It is also very extensible (reuse of existing ontologies, export in various formats, visualization plug-ins, etc.). Finally, the use of inference engines is very easy. These engines allows automatic reasoning on our ontology, and we used some of them (see 5).

4. Equations

In a former work on accounting standards [Teller (2006)], we proposed the following general formula for all the accounting standards in IFRS:

$$\begin{aligned} Standard &= Prin \cup \sum_i Def_i \cup \sum_i Proc_i \cup \sum_i O_i \\ &\cup \sum_i S_i \cup \sum_i C_i \cup \sum_i Doc_i \end{aligned}$$

where Prin is a set of accounting principles, Def is a definition, Proc a procedure, O an obligation, S a suggestion and Doc a documentation to provide. In other words, we divided the standards in general categories of actions. That allowed us to compare and to classify them. The next step of the modeling process is to describe each component of this formula, for each standard. Once this step is achieved, it becomes possible to use inference mechanisms.

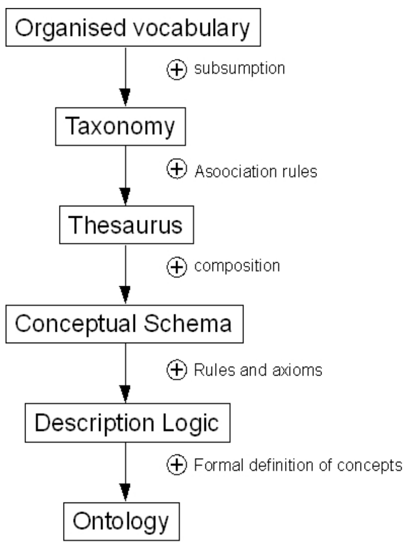

Figure 4. The origin of ontologies

4.1. *The Elements of the Modeling Language*

The first step is to constitute a set of constructors. The choice of the constructors conditions the expressiveness of our language. The resulting set of constructors is represented in table 1. Some of them are classical, and their semantic is obvious: := stands for a definition of a new concept or relation, □ for object disjunction, ∩ for object conjunction, □ is the conditional operator. We also use quantifiers: □ for existential quantification and □ for universal quantification. ≥ n and ≤ n are cardinality restriction operators. All these elements are quite familiar.

But we had to represent more specific operators because of accounting standards particularities. Some accounting standards offer to choose between several possibilities, but indicate clearly a preferential one (and often, choosing another implies to mention and to justify this choice). We will represent this situation with the operator, and the preferential choice will always be on the left side. We also introduced comparison operators (< and >) in order to compare concepts. The nature of the comparison depends on the nature of the concepts: we are comparing the valuation of concepts, which can be a duration, a cardinality, a monetary amount, etc. So, all concepts are not comparable together (the categories of comparable concepts can be found in our description of accounting ontology in chapter 5). The operator is very typical of the accounting standards. It stands for the cases where a standard indicates that the user has to set if two concepts are comparable, but the method for the comparison is not specified. It is the responsibility of the user to determine what is the criterion for the concepts for being comparable, depending on the context. The operators min and max imply to select between a set of concepts, the criterion for the selection depending on the context and the nature of the related concepts.

The most particular constructors are those inspired by modal logic. We choose to introduce them in order to capture the semantic of some elements of the accounting standards for which first order logic operators would not be adapted. These operators are: mandatory, possible, probable and major. For a detailed description of the semantic of these operators, see [McCarthy (1996)].

Table 1. The constructors of the modeling language

$\cup, \cap, \exists, \forall, \geq n$ $\leq n, \neg, \subseteq, \Rightarrow, :=$	classical operators
$\sqcup$	disjunction with a preferential choice
$\simeq <>$	comparable, superior, inferior
min, max	selection operators
OB, PO PR, MAJ	modal logic operators (mandatory, possible, probable, major)

Then, we extracted a set of primitive concepts and relations from the conceptual framework of accounting and from IFRS definition texts. These primitive objects are the elements from which all other objects can be constructed. Some of the top-level objects in the subsumption hierarchy are represented on table 2, and also some primitives that compounds some top-level elements in the subsumption hierarchy.

With these elements, we are now able to represent composite accounting elements. For instance, the following example gives the definition of an asset: it is an economic good that the entity controls and from which it will obtain economic advantage.

$$asset := good \cap \exists economic\ entity.(control \cap economic\ advantage)$$

Recursively, we can represent lower-level objects in the subsumption hierarchy. Thus, once we have the definition for asset and long-term (which would be: *long-term := period > fiscal year*), we can define the following composite concept :

$$fixed\ asset := asset \cap material \cap$$
$$(economic\ good).(useful\ life).long - term \cap$$
$$entity.(economic\ advantage).measurable$$

Table 2. Primitive concepts and relations

primitive concepts	primitive relations
economic entity	estimation
transaction	measurable
economic good	book recording
financial statement	control
period	economic advantage
fiscal year	recognition
account	significant
risk	forecasting
user	raising / diminution
accounting principles	material
performance	financial
	cost

financial statement	period	accounting principles
balance sheet	generic period	accruals basis
profit & loss	fiscal year	going concern
notes	useful life	consistency
stockholders equity statement		pertinence
cash flow statement		materiality
		reliability
		true and fair view
		substance over form
		neutrality
		completeness
		comparability
		prudence

4.2. *Extraction of Rules*

Once the objects have been defined, it is possible to represent any accounting statement. The example in figure 5 represents how to record a loss resulting from a long term contract. The resulting rule would be:

$$
\begin{gathered}
(loss.reliable \Rightarrow loss.provision) \cup \\
(\neg loss.reliable \Rightarrow ((loss.estimation.reasonable \Rightarrow \\
(loss.estimation.likely.provision) \cup \\
(loss.estimation.minimum \cup notes.mention))) \cup \\
(\neg loss.estimation.reasonable \Rightarrow \exists notes.mention)))
\end{gathered}
$$

Let us see a simple example of inference which can be made with this language. If an asset is an inventory item, its depreciation is made by provision, and if it is a fixed asset, its depreciation is made by amortizing. As a consequence, if an asset has amortizing, we can conclude that it is not an inventory item. The representation of this simple case with our language is:

$$\begin{cases} \textit{fixed asset} \Rightarrow \textit{asset} \cup \\ \qquad (\textit{depreciation} = \textit{amortizing}) \\ \textit{inventory item} \Rightarrow \textit{asset} \cup \\ \qquad (\textit{depreciation} = \textit{provision}) \end{cases}$$

$$\textit{asset} \cap \exists\, \textit{amortizing} \Rightarrow \neg\, \textit{inventory item}$$

We have now a language for achieving our goal, creating a formal representation of accounting standards. Indeed, each accounting standard can be translated in one or several constructors like in the example above.

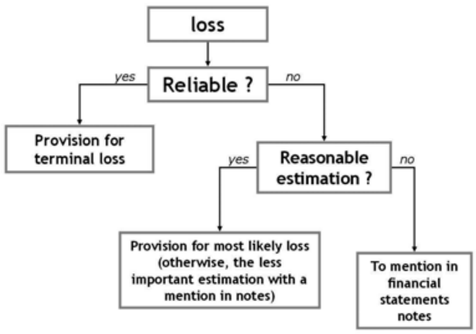

Figure 5. The accounting statement: recording a loss on a long term contract

4.3. *Results*

We extracted the rules from the whole corpus of documents of the accounting standards IFRS. Of course, we faced classical modeling problems (determining whether an element is best represented by a concept or a relation is not always obvious, knowing whether the introduction of a new constructor is appropriate can be a problem, etc.). We also had problems related to the nature of our corpus : the natural language. The modeling of this material is complicated [Aussenac-Gilles (2000)] [Bateman (1993)]. Representing some very specific measures would have required the introduction of specific elements in our language, and would have increased the complexity of our language. In some of these cases, we decided not to represent the problematic feature.

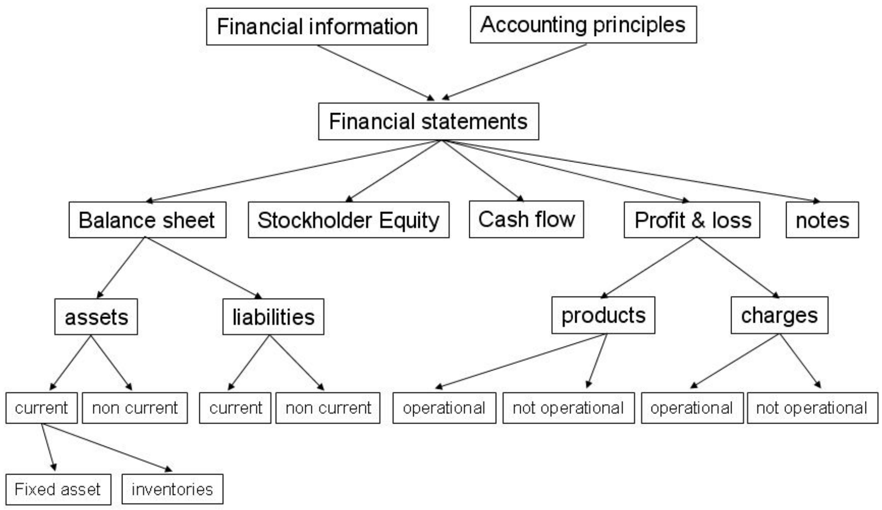

Figure 6. A part of the subsumption hierarchy

Once this task is completed, a subsumption hierarchy is available. We represented in figure 6 a part of the subsumption hierarchy of accounting standards. It shows that we obtain a result that is consistent with the general representations of financial statements, as all accountants use them.

5. The Semantic Model of Accounting Standards

5.1. *Motivations*

In this part, the goal is to describe the accounting standards from a semantic point of view. We will focus on the meaning more than on the framework. We have seen that the syntactic description led us to build a language which contains the key-components of accounting, and these key-components will be used to describe how the accounting standards work. The components of the language were not created for the occasion, they were chosen among all the notions commonly accepted and used by the stakeholders of financial information. That's why it is important to define what meaning we attributed to each component, in order to avoid ambiguities.

There may be several sources of mistakes. We might find synonymies, when several terms are used for the same component or the same action, or conversely we might find a polysemia when a term can have several meanings.

Let us illustrate this by our fixed assets example. The notion of ”spend” or ”cost” is used for a charge, in other words something the entity will have to pay and which affects its result, but it can also be used for an investment that will be recorded as a fixed asset and thus will not affect the result. The difference is very important, and the financial audit takes a special care to the correct distribution of the costs between charges and fixed assets. Our model must be able to deal with this difference.

There’s another possible confusion when several interpretations are contradictory. In this case, we must be able to recognize which interpretation is considered. In our example, the procedures described for the fixed assets treatment are not always coherent. According to IAS 16, the initial recording must be made using the historical cost of the asset, recording at the fair value is an optional treatment that must be justified. But in IAS 40, concerning the investment property, which is particular cases of fixed assets, the recommended treatment is just the opposite. To face this problem, it’s said in IAS 16 that the standard must be applied to all fixed assets unless another standard exists for the particular case considered; nevertheless this demonstrates the need for our model to manage priority between standards when their application domain overlap.

But the more frequent case is when the meaning of a notion depends on the context. With the IFRS, this case is even more frequent than before. The nature of a concept might change because of different management choices, because its treatment changes with the activity areas, with the country, etc. For instance, the costs for investment property are recorded:

$$\begin{cases} \text{in stock, if the entity's activity is real estate;} \\ \text{as fixed assets otherwise.} \end{cases}$$

Our semantic model must be able to face all these problems, because all the terms used in the syntactic model will belong to this thesaurus, and they must be clearly defined. Another point is to consider the relations and dependencies that might exist between concepts [Mirbel (1995)]. Every action impacting an element is likely to impact others. This implies the building of a dependency graph.

5.2. *The Ontology*

The final objective is to build an ontology of accounting notions, representing the whole knowledge of the domain. As for the syntactic description, it does not pretend to be the only possible ontology, the objective is to be adapted to our need which is to give financial information modern structures [Woodroof (2001)].

We built our ontology in a bottom-up way. We made this choice because our corpus is close: we had a finite set of document in which it was relatively easy to collect all meaningful terms. We then grouped them in categories, and repeated this operation until the top-level. Each step was validated by some experts of the domain in the partner company Ernst & Young, which is a guarantee that our result is coherent with usual habits of accounting professionals. We already noted in previous chapters that the hierarchy of the ontology and the subsumption hierarchy of concepts of the modeling language are similar (it is more detailed in the lower levels in the case of the ontology, but the upper levels are similar).

We used the software Protégé, from Stanford University, to represent our ontology. This software allows graphical representation and modeling, and has the ability to convert and export the ontology in various format, including XML. This choice is therefore appropriate in our case. Figure 7 is a screenshot of it.

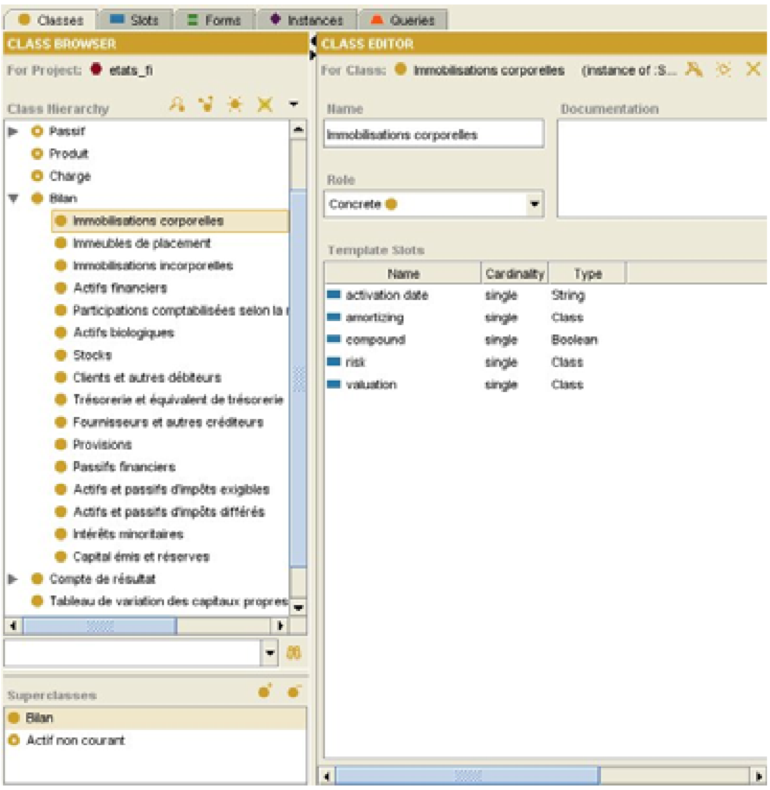

Figure 7. A screenshot of Protégé

We can see also on figure 7 the particular case of the element Property, Plant and equipment: it is a subclass of Balance sheet and non current asset. We can see on the lower-right part of the screen the slots (or attributes) of this element. These slots have a particular type and a cardinality, and can be an instance of another class of the ontology (in our example, it is the case for the slots risk, amortizing and valuation).

In order to be able to use inference engines (see 5.3), we stored our ontology in the OWL format. This language is the W3C recommendation for web ontologies. It is becoming the standard language to store this kind of objects, and it is easier to reuse other ontologies if they are all written with the same language [McGuiness (2004)]. In fact, we built our ontology with the objective to stay compliant with a subset of OWL, called OWL-DL. This subset guarantees that the resulting ontology has a limited complexity, so that inference engines will be able to process that ontology in finite time.

5.3. *Results*

Our ontology was primarily built from IFRS standards definition texts. Of course, it has some influence on the resulting ontology. IFRS standards focus on the importance of the information to be given to financial markets and investors. In other words, these standards are reporting-oriented with a strong financial connotation, and that is necessarily the case of our ontology too. We could imagine other possibilities, a management-oriented ontology for instance, which would also be an ontology of accounting, but where the definitions of terms, the relations between them and the corresponding hierarchy would be considered from the point of view of someone in charge of the entity that produces the financial informations. In our case, we adopted an external point of view, which could correspond for instance to analysts, investors or regulators: the difference may be important. There is no point in telling that one of these ontologies is better than the other, it depends on what are the applications of the ontology. But in our case, we did not really have the choice, because of the nature of the documents we used. Building our ontology from another point of view would have required to work on different documents and sources. Figure 8 represents a piece of the resulting subsumption hierarchy.

The consistency of our ontology was confirmed by :

- the analysis of experts of this domain,
- the coherency with the results of the syntactic representation,
- the use of inference engines.

We used two different inference engines : Racer [Haarslev (2001)] and Pellet. These engines, among other verifications, check that the internal structure of the ontology does not contain any contradiction or impossibility. They are a good proof that our ontology is well-formed. The validation by the experts of the domain, on the other hand, ensures that the ontology is consistent with the expectations of the professionals of the domain.

6. The Applications of our Model

Once the model is built, we can use it for various tasks related to financial information manipulation. Such a model is especially useful in order to find, to extract and to analyse some particular informations automatically within a large amount of data, and therefore it requires a numeric standard to store the information. We have seen earlier that XBRL is a perfect candidate for that.

Some of the applications we built were inspired by the main task of the company who cooperated on this work, the financial audit. Others are more general, and concern other financial data manipulators, like regulators, investors, banks, analysts, etc.

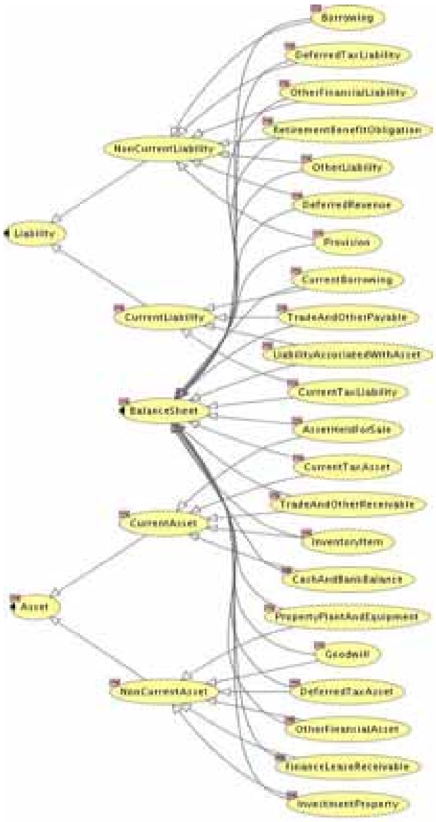

Figure 8. A piece of the resulting subsumption hierarchy

6.1. *Detection of Typical Dangerous Situations*

The first step of almost every financial audit consists in an overall assessment of the current financial statements. The goal is to obtain a global understanding of the economic situation of the audited company. Indeed, it is possible to infer a lot of information that are not explicitly written in the financial statements, and it is done basically by comparing the current situation with other well-known and typical situations. This task is very important, since the result of this preliminary analysis has a great influence on the way the audit mission will be lead.

This work is still done in an implicit manner. We wanted to demonstrate that our model can help to formalize this task, and thus to share the results and to homogenize the results.

6.1.1 Methodology

This work involves several step :

- establishing a list of significant criterions,
- express these criterions using our modeling language,
- setting the eligible values for each criterion,
- balance the relative importance of the criterions,
- find the thresholds to interpret the resulting scores.

The first step requires the help of the experts of the domain. They tell us what are the significant facts which would represent clues for our evaluation. The second step on the other hand is to be done by the modeler. He has to find the right logical expression that will represent the situation described by a criterion, using the modeling language. One must be able to evaluate these expressions, in other words to give marks representing to what extent the current situation matches the criterion. The third step is a direct consequence of the second : we had to set up what are the possible marks that could be given when evaluating a criterion. Usually, the marks are taken in the set {-2,-1,0,1,2}, but the set can be different, if it is required in a particular situation. Negative values stand for the cases where the evaluation contradict the situation that we are trying to identify. Indeed, a negative mark will lower the risk in the overall estimation. The mark 0 means that the criterion brings no information in this particular case. It is possible to admit marks higher than 2 for very significant situations, for instance if the evaluation gives a very surprising result.

The goal of the last two steps is to take into consideration the links that might exist between the criterions. First of all, all the criterions have a different degree of significance. Moreover, there might exist mutual influences, a particular evaluation of a criterion may emphasize the evaluation of another criterion, or on the contrary decrease this evaluation. This is the reason why we gave a weight to each criterion. The overall estimation of the situation consists then on the weighted sum of the marks obtained by all the criterions related to this situation (see figure 9).

There is an important property concerning the weights. In order to compare several experimentations of our methodology, the sum of all weights has to be the same. For practical purposes, we let the experts set all the weights without any constraints, and then we divided them by the sum of all of them. It guarantees that the sum of all weight is always equals to 1, whatever situation (and whatever experts) is considered. In our example, we should divide every weight by 1.6.

This weighted sum can then be analysed and compared to others by the experts of the domain. This work was already done in every audit mission, but without formal basis. This methodology allows a better sharing of the results, and guarantees that all the concerned criterions were taken into account.

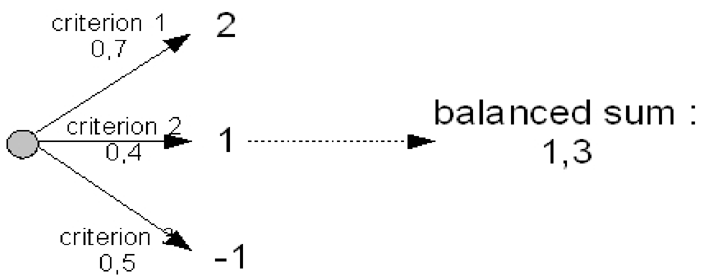

Figure 9. The weighted sum of three criterions, before normalization

6.1.2. Case Study

We applied this methodology to real situations, with the help of the company Ernst & Young. The experts were the auditors of this firm. In this paper, we will just give a brief overview of the results of this work, for a more detailed description, see [Teller (2007)].

The situation described here is the risk of overestimation of the result, which is one of the most common risk financial auditors has to detect. Here are the criterions for this particular situation :

- fixed assets amortized on a longer period,
- stock estimation method changed since last fiscal year,
- stock appraisal raised since last fiscal year,
- recording of potential capital gains,
- important expenses transferred as fixed assets,
- change of doubtful debts recording method,
- selling of fixed assets followed immediately followed by a leasing contract of these assets.

These criterions represent accounting procedures that lead to a raising of the result of the current fiscal year. These procedures must respect some conditions to be legal, and if too many of them are doubtful, the expert concludes that the audited company tries to overestimate the result.

Here is the representation of these criterions using our modeling language :

$$\begin{aligned}&\exists(Asset_i \subseteq FixedAssets)\\&Exercice1.Asset_i.LifePeriod <\\&Exercice2.Asset_i.LifePeriod\end{aligned} \tag{1}$$

$$\begin{aligned}&\exists Stock_i \subseteq Stock\\&(Exercice1.Stock.Appreciation.AccountingMethod \neq\\&Exercice2.Stock.Appreciation.AccountingMethod)\\&\cap (Exercice1.appraisal(Stock_i) <\\&Exercice2.appraisal(Stock_i))\end{aligned} \tag{2}$$

$$\begin{aligned}&\exists Stock_i \subseteq Stock\\&Exercice1.net\_value(Stock_i) <\\&Exercice2.net\_value(Stock_i)\end{aligned} \tag{3}$$

$$\begin{aligned}&\exists FinancialAsset_i \subseteq FinancialAsset\\&FinancialAsset_i.recording(CapitalGain) \cap\\&\neg FinancialAsset_i.Outing\end{aligned} \tag{4}$$

$$MAJ\ (Asset.recording(ExpenseTransfert)) \tag{5}$$

$$\begin{aligned}&\exists DoubtfulDebt_i \subseteq DoubtfulDebt\\&(Exercice1.DoubtfulDebt_i.Appreciation.AccountingMethod \neq\\&Exercice2.DoubtfulDebt_i.Appreciation.AccountingMethod)\\&\cap (Exercice1.appraisal(DoubtfulDebt_i) >\\&Exercice2.appraisal(DoubtfulDebt_i))\end{aligned} \tag{6}$$

$$\begin{aligned}&\exists(Entity_i \subseteq FinancialAgency, Contract_i \subseteq Contract)\\&Contract_i.clause(Entity_i.assurer(\forall DoubtfulDebt))\end{aligned} \tag{7}$$

$$\begin{aligned}&\exists Asset_i \subseteq FixedAssets, Contract_i \subseteq Contract\\&Exercice.selling(Asset_i) \cap\\&Contract_i.clause(Entity.lease(Asset_i))\end{aligned} \tag{8}$$

As we mentioned before, we had to set up the possible marks for the criterion's evaluation, and a weight for each criterion. The weights were then normalized, in order that their sum was equal to 1. The possible marks and the weights for each criterion are represented in table 3.

This procedure demonstrates the possibilities of our model in the case of a financial audit. It makes possible to communicate formal and precise instructions to the expert in charge of a mission, and in return, this expert can communicate the results and compare them with other results, in an easier way than it is done until now.

We will now illustrate that similar procedures can be useful for other accounting professionals.

criterion	marks	weight	balanced weight
1	{0,1}	0,5	0,11
2	{-1,0,1}	0,7	0,15
3	{0,1}	0,5	0,11
4	{0,1}	0,3	0,06
5	{0,1,2}	1	0,21
6	{-1,0,1}	0,8	0,17
7	{0,1}	0,5	0,11
8	{0,1,2}	0,4	0,09

Table 3. The marks and the weights for this procedure

6.2. *Discriminatory Analysis*

Regulators are another family of big financial data consumers. One of their missions is to collect and to process the biggest amount of informations within the shortest time period. The bigger is the amount of data taken into account, the better is the analysis produced by the regulators. This analysis aims at being a guaranty of quality of the publicly published financial informations [Demolli (1992)]. It is easy to understand that the improvement of financial data manipulation tools is a crucial issue for them, as the financial publications are more and more voluminous and frequent [Johnson (1992)].

One of the task of the regulators is called Discriminatory Analysis. It consists in the calculation of one or more marks, for a single company or for several together. This calculation is based on classical financial indicators, extracted from the informations published. The obtained marks are then used to separate the companies into subsets : for instance, a mark under a given threshold would be a clue for a potential bankruptcy [Altman (1968)] [Taffler (1983)] [Ghesquiere (1983)].

The marks are computed using a linear combination :

$$Z = a1\ X1 + a2\ X2 + \cdots + an\ Xn$$

where Z is the overall synthesis variable, **ai** are discriminatory coefficients, and **Xi** are independent variables extracted from the published informations.

We will now investigate how our model of accounting standards can improve this analysis. First of all, it is possible to represent the different criterions of the formula (the Xi) by logical expressions, and as a consequence, it becomes possible to use the automatic inference mechanisms offered by this kind of formalism. At this point, the procedure is similar to the procedure described in 6.1.

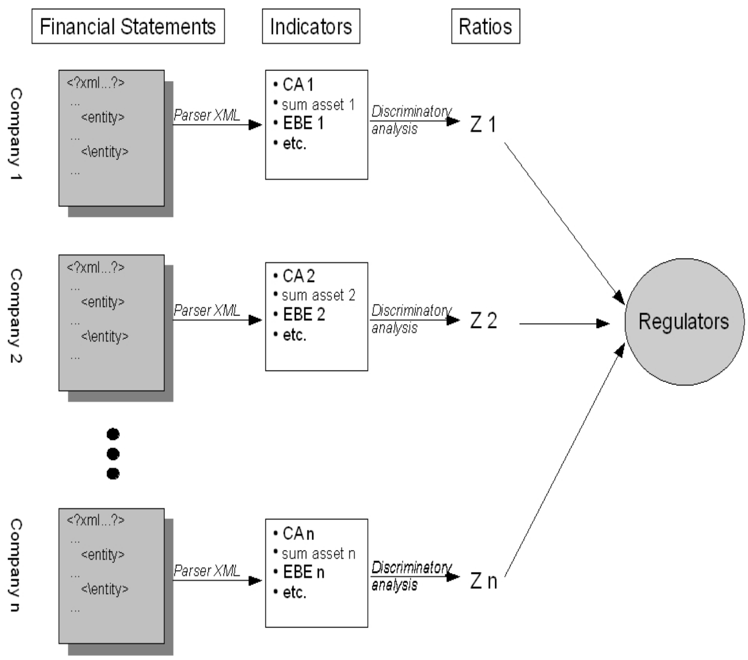

Figure 10. The different steps for an automatic analysis

But the potential benefits of this methodology are improved if the sources of information are correctly formatted, for instance as XBRL instance documents. A simple parser allows an automatic extraction of data, and makes the evaluation of the Xi automatic. In these conditions, it becomes possible to process the informations as soon as they are published, and the resulting analysis is greatly improved. Figure 10 shows how these tools work together.

6.3. *Other Possibilities*

There are other possible applications for our model. In fact, in almost all the cases where financial information has to be processed, a formal model of the contained knowledge is useful. Let us consider rating agencies. Their subjectivity was pointed during every big financial scandals [MasterFinance (2007)]. This subjectivity is the result of several factors, as shown in figure 11 :

- each agency has only a partial view of the global economic environment,
- they are limited to the informations publicly published,
- each agency has an influence on the other agencies, and conversely.

Our goal is not to give details about this subjectivity, and not either to suggest solutions. We just want to illustrate that rating is another possible field where our model can be a useful tool, if it is used in applications similar to those described in 6.2.

Indeed, the possible uses of our model are multiple. For instance, financial auditors have to check that the document attested to be correct contains exactly the same data as the one that have been audited. Till now, this task is still done manually. Automatic data extraction tools would be very useful, and would allow large time savings. There is a lot of examples like this one, and that is why we are convinced that our model can be employed in many applications.

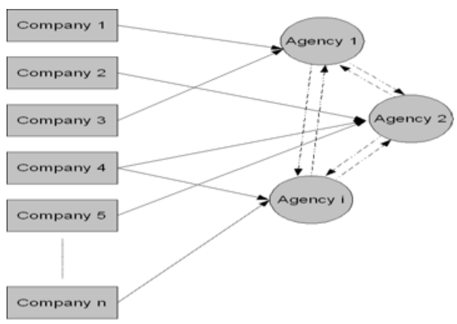

Figure 11. The factors of subjectivity of rating agencies

7. Conclusion

Our goal is to build a formal representation of accounting standards. This leads to an ontology of accounting notions and a general schema of accounting standards. We want to give to financial information a formalism adapted to new technologies. The accounting is still based on the principles that emerged in fourteenth century, which were adapted to the support of information available in this period, writing and bookkeeping. But digitalization offers a lot of possibilities if financial information is able to adapt. That is why the issues of the formalization are so important, because in spite of the inertias, the change will necessarily occurs. The schema we are proposing tries to make this evolution easier.

The main result of our work is a model of representation of international accounting standards (the IFRS). This model is composed of a syntactic part and a semantic parts. The syntactic representation uses a formalism inspired by description logics. This formalism allows the use of non monotonic logical operators, and these operators were very useful to represent some of the most typical features of accounting standards. The semantic of accounting standards is represented using an ontology, because this tool is very convenient for our purpose, and it allows automatic knowledge-oriented reasoning, like classification within a taxonomy. This ontology is used to fix the meaning of the concepts used in our model.

The possibilities of our model are illustrated by practical case studies. These cases show that the availability of this model may greatly improve some tasks related to financial data manipulation. A possible further work could be to establish a similar model for other accounting standards, for instance the USGAAP. If the same methodology is used, the two models could be easily compared, ans that would be a great help in order to represent the differences between these two referentials.

References

Claude Alia and Robert Descargues. Modélisation et comptabilité. in *Encyclopedie de comptabilité, contrôle de gestion et audit*, Editions Economica, 2000.

Edwards I. Altman. Financial ratios, discriminant analysis and corporate bankruptcy. *Journal of Finance*, 1968.

Nathalie Aussenac-Gilles, Brigitte Biebow, and Sylvie Szulman. Revisiting ontology design: A method based on corpus analysis. *Lecture Notes in Computer Science*, 1937:172–188, January 2000.

John A. Bateman. Ontology construction and natural language. In *Proceedings of the International Workshop on Formal Ontology*, pages 83 – 93, Padova, Italy, March 1993.LABSEB-CNR.

ISACA Standards Board. Continuous auditing : Is it fantasy or reality? *Information Systems Control Journal,* vol 5, 2002.

Daniel Boussard. La Modélisation Comptable en Question(s). *Economica*, 1997.
Jose Cardoso and Ana L. Nunes Escorcio. Editing tools for ontology construction, 2007.
Jean-Guy Degos. La révolution copernicienne de la comptabilité reste à faire.
Eric Demolli. *Vers une méthodologie de l'Audit des systèmes d'information comptables.* PhD thesis, Université de Nice, Département des Sciences de Gestion, 1992.
Fabien Gandon. Ontologies informatiques. Technical report, INRIA Sophia-Antipolis, team ACACIA, 2006.
James Garson. Modal logic. Stanford Encyclopedia of Philosophy, 2007.
Pierre Gensse. Modèle comptable français. in *Encyclopedie de comptabilite,controle de gestion et audit*, Editions Economica, 2000.
Sylvie Ghesquères and Bernard Micha. Analyse des défaillances d'entreprises. *Centrale des bilans de la Banque de France*, 1983.
Volker Haarslev and Ralf Moller. RACER system description. *Lecture Notes in Computer Science,* 2001.
David Henry. The growing revolt against the sec. Business Week, april 23, 2007, 2007.
Institut Canadien des Comptables Agréés. Incidences du langage xbrl sur la vérification et le contrôle. *Dossier d'Information du Comité Consultatif sur les Technologies de l'Information*, May 2002.
O. Johnson. Business judgment versus audit judgment: why the legal distinction. *Accounting Organization and Society*, 17:205 et sq., 1992.
David Jouve. Modélisation Sémantique de la Réglementation. *PhD thesis, Informatique et Information pour la Société*, November 2003.
R.W. van Kralingen. Frame-based Conceptual Models of Statute Law. Kluwer Law Intl, 1995.
Master Finance Belgium. Subprime : de la responsabilité des agences de notations., 2007.
John McCarthy. Modal logic. Some Philosophical Problems, 1996.
Deborah L. McGuinness and Frank van Harmelen. Owl web ontology language overview. http://www.w3.org/TR/2004/REC-owl-features-20040210/, 2004.
Microsoft Corp. Improving financial reporting and analysis using xbrl and the microsoft office system. Technical report, Microsoft, 2005.
Isabelle Mirbel. Semantic Integration of Conceptual Schemas. In Proceedings of the First International Workshop on Applications of Natural Language to Databases (NLDB'95), pages 57–70, Versailles, France, 1995.
Daniele Nardi and Ronald J. Brachman. An introduction to description logics. pages 1–40, 2003.
Natalya Fridman Noy, Michael Sintek, Stefan Decker, Monica Crubezy, Ray W. Fergerson, and Mark A. Musen. Creating semantic web contents with protégé-2000. *IEEE Intelligent Systems*, 16(2):60–71, 2001.
Jim Richards. The anatomy of an xbrl taxonomy. Technical report, *Murdoch University,* Australia, May 2002.
Jim Richards. Tha anatomy of xbrl instance documents. Technical report, *Murdoch University*, Australia, May 2002.

Richard J. Taffler. The z-score approach to measuring company solvency. *The Account's Magazine*, 1983.

Pierre Teller. Representation of accounting standards: Creating an ontology for financial reporting. In *Walter Dosch and William Perrizo, editors, SEDE*, pages 234–239. ISCA, 2006.

Pierre Teller. Une Représentation Formelle des Normes Comptables. *PhD thesis,* Université de Nice Sophia-Antipolis, 2007.

Nicolas Véron. L'Information Financière en Crise. *Odile Jacob*, 2004.

Jon Woodroof and DeWayne Searcy. Continuous audit : Model development and implementation within a debt covenant compliance domain. *International Journal of Accounting Information Systems*, 2001.

Stanley Zarowin. A napster for financial data? *AICPA*, 2003.