AN ONTOLOGY-BASED SEMANTICS FOR INTERBUSINESS PORTALS

Oualida BOUZID*
Nacereddine ZAROUR**

*Faculty of Engineering, University Larbi Ben M'hidi of Oum-El-bouaghi, Algeria
**LIRE Laboratory, University Mentouri of Constantine, Algeria

bouzid_oualida@yahoo.fr, nasro-zarour@umc.edu.dz

Abstract

In this work, we propose a solution to solve the problem of ambiguity in interbusiness portals. We define by ambiguity, any type of terminological variation defining the same object or aspect, that can be due to an ambiguity related on the language (synonymy and polysemy), or to the difference in terminological choice of the various companies. To minimize the appearance of this problem with an aim of improving collaboration of the companies, we propose a solution based on a semantic homogenisation of the terminology used on the portal. This semantic homogenisation would be concretized by the development of ontology.

Keywords: *knowledge portal, semantic ambiguity, ontology, and equivalence.*

1. Introduction

An interbusiness portal, particularly a knowledge portal is a plat-form of exchange and sharing information for a collaborative work. The success of this collaboration depends primarily on the shared knowledge, which must be comprehensible by all the information systems (human agent and software application) of the various companies concerned with this portal. Therefore it is clear, that to improve this collaboration and to ensure a good development, it is necessary to improve the access to the knowledge and to allow a good interpretation of the information present in this kind of portal.

To improve the business cooperation, the various problems hampering the diffusion and the correct interpretation of information in a knowledge portal have to be fixed. Some works try to improve these portals. Two of the greatest project in this field are OntoWeb [4] and HARMONISE [8]. They aim at curing the semantic and interoperability problems without aiming at a precise problem.

The problem we want to solve is that of ambiguity when several businesses share information. There exists linguistic variation in the terminology used by these businesses using the same language but different terms.

To limit the negative effects of this problem and improve the collaboration between businesses, we suggest a solution based on a semantic homogenization of the terminology used on the portal. This semantic homogenization is implemented through the elaboration of ontology.

The ontology is an independent layer that the existing portal will use. This portal is assumed to have a specific theme, and is mainly informational.

Initially, we propose a process for the elaboration of such ontology. This process is based on the equivalence relations. For the elaboration of this ontology, we have adapted Methontology [7] to our needs. The choice of the methodology for ontology development is a very important step, on which depend both the quality of the ontology and the amount of effort needed to achieve the goal. The suggested process is illustrated by a case study which is implemented using OWL FULL and PROTEGE, and validated using the RACER tool.

The rest of the paper is organized as follows. In the following section, we present some related work to the ontology issue in portals. In section 3, we develop the suggested ontology for interbusiness portals and detail its characteristics. We show the ambiguity problem consequences on portals in section 4. In section 5, we give a solution to this ambiguity problem. In section 6, we illustrate the proposed ontology development process on a real portal. Section 7 summarizes the contributions of this paper and the research perspectives.

2. Related Work

A lot of works have dealt the ontology issue in portals. We present those that are similar to the proposed work.

"OntoWeb: Ontology-based information exchange for knowledge management and electronic commerce" [4] aims at bringing together researcher and industrials to enable the full power of ontologies. The project aims at improving information exchange in some areas such as information retrieval, knowledge management, electronic commerce, and bio-informatics. In essence, knowledge portals exploit ontologies for achieving a conceptual foundation for all functionalities that are offered by the portal. The SEAL framework for developing and managing knowledge portals exploits Semantic Web technologies to offer mechanisms for acquiring, structuring, integrating, sharing and accessing distributed knowledge between human and/or software agents. Up to now, SEAL puts emphasis on supporting the acquiring and structuring of knowledge by semantic annotation, the automatic generation of navigational views, mixed ontology, and content-based presentation. SEAL uses Methontology but does not consider semantic ambiguity in all its aspects. It districted between terms, symbols, and concepts. This distinction does not completely solve the problem of ambiguity, which is solved in our approach.

Harmonise [8], is another example for project in the semantic portal field. It uses ontology as much as semantic mediators in the community portal. In order to improve interoperability, this approach uses ontology to harmonize information shared between diverse applications. Contrary to our work, we are not interested in applicative interoperability but in informational interoperability.

3. An Ontology for Interbusiness Portals

One solution to improve collaboration between businesses is to develop ontology. Indeed, collaboration between businesses may take place through information exchange (file, article, etc.). In this work, we consider inter-business collaboration taking place through information sharing across a common portal.

In information sharing environment, such as an interbusiness collaboration portal is difficult, and possibly useless, to normalize the terminology used by the various businesses involved so as achieving better collaboration. Instead of requiring each business to use the same terminology, a possibility is to look for a way to make them understand the shared information. This is what is meant by using ontology for semantic homogenization of the information shared on a portal. Harmonizing the terminology is understood as homogenizing the semantics of the terms used. The use of ontologies for information portals is not a new field; several such projects exist, such as SEAL [4].

In this work, our interest is confined to a single aspect, namely developing ontology capable of eliminating or minimizing ambiguity and highlighting the equivalence relations between the terms. The developed ontology would be an additional, independent layer to be invoked by the content management system of the existing interbusiness portal [3]. As shown in figure 1, the information exchange between businesses members, transiting by the interbusiness portal, would call upon the ontology.

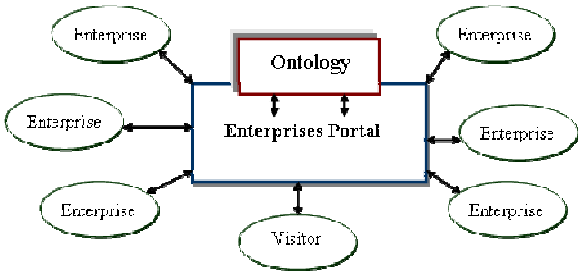

Figure 1: Ontology as much as semantic mediator

In practice, developing ontology for a field requires a lot of effort, especially for acquiring specialized knowledge and establishing semantic links (classes and relations). Towards this end, it is necessary to analyze real needs. Developing ontology is undertaken only if there is a real ambiguity problem hampering cooperation, or if there is a real need for having a notion of meaning in order to improve search results relevance [11].

Still, each business has a specific terminology managed through its information system [9]. Normalizing and standardizing information-carrying terms, required for collaboration, entails redesigning the whole information system for each business. To avoid this, we propose the development of an ontology above of the portal which would ensure harmonization of the terminology used and would have a role similar to a translator base.

For this work, we assume that the portal already exits, and the ontology is developed as a super-class (fig. 2). Using the ontology will improve access to the information, and will make it more understandable by all users, human agents of the different businesses and software agents developed independently.

In this work, we are interested in interbusiness collaboration portals that are based on

information exchange. This is a space which gathers in a single interface contents and services coming from the different information systems for different businesses and revolving around a theme.

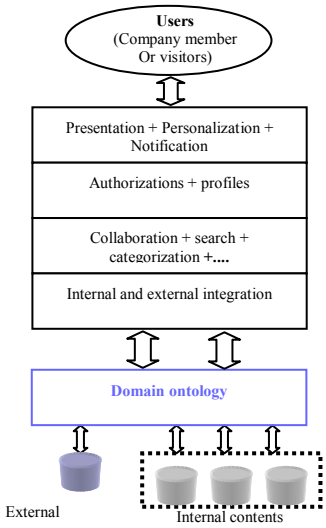

Figure2 : Considering an ontology in a Portal Functional Architecture

This theme will represent the field of the ontology we intend to develop in order to normalize the terminology used on the portal and thus solve the ambiguity problems. To develop such ontology, it is necessary to define the "fields of information" of the portal, its main theme and the sub-themes it covers. Then, on this basis, build a corpus from which the knowledge to model is extracted. The arborescence of the portal has to be analysed, as well as its functions and its goals, so as to extract all the information to be used in building the corpus.

We are interested mainly in two kinds of content, according to the type of exchange existing between member businesses: editorial content (article, dispatch, file, etc.), and marketing content (product catalog). In the chosen methodology for developing such ontology, Methontology, the task of knowledge acquisition will allow the identification of the terms to include in the ontology.

4. The Ambiguity Problem in Inter-Business Portals

In this work, we are mainly interested in information-management collaborative inter-business portals. Inter-business portals represent an exchange and collaboration environment where several businesses share information (B2B portal). The success of such collaboration depends primarily on the information shared. This information must be understandable by all the information systems of the different businesses involved in this portal.

One of the primary functions of an inter-business portal is to provide access to the set of contents, services, and internal and external applications of the different businesses. The available information may be shown in several space parts of the portal:

- A community space: individuals and groups can communicate directly (the notion of group or community is very clearly expressed).
- A services' space: search, email box, office services, etc.
- An informational searchable space: giving users access to internal and external information, with the ability to customize this information according to the user's profile presented to each user according to its profile, information tailored to the user's needs.

The access to this content, which is available in interbusiness portals, is not always possible, unsuccessful search, unintelligible or misinterpreted information, etc. This is linked to a certain number of problems, among which is ambiguity. Terminological ambiguity is related to some aspects of natural languages such as synonymy or polysemy, as well as to the specificity of the terminological choices to each business. This problem lowers search efficiency and information identification: the information exists on the portal, but some businesses, because they use different terms from those of the publishing business, cannot access it.

Ambiguity can result from:

Language specificities, such as:

- Homonymy: when two semantically distinct terms have the same pronunciation. In this study, this aspect is ignored as we are mainly interested in written documents.
- Synonymy: several formally distinct terms have the same meaning, resulting in several signs for the same object. *The semantic relation that holds between two words that can (in a given context) expresses the same meaning*.
- Polysemy: the same sign has several meanings. *The ambiguity of an individual word or phrase that can be used (in different contexts) to express two or more different meanings.*

The specific terminological choice of the business may result in terminological variation that leads to semantic ambiguity. The company may use specific terms to identify a specific object or product. Thus, if each company in a specific field uses a company-specific, distinct term to refer to an object (aspect, product, etc.), semantic confusion is a possible result.
Another problem similar to those of ambiguity and which is considered in this work is that of equivalence. While ambiguity results from several terms pointing to the same object, equivalence problems come from distinct concepts that present some equivalence on a specific aspect (usually, functional equivalence, e.g. in the field of transportation, "car" and "bus" fulfil the same function, and they are also equivalent in that they are both "public transportation").
The company can have narrower terms, with certain object or certain product. Therefore if each company working in a field uses a narrower term and clean with it to describe an object (aspect, produced, etc.), there could be a confusion.

5. A Solution to the Ambiguity Problem

We have presented the ambiguity problem as one related to understanding the meaning of a given concept or as the existence of two forms with the same meaning. The solution we propose is to establish a link that implements the semantic equivalence between the two forms. For each form, we will identify all the equivalent, or semantically similar forms. Then we define an equivalence relation that links all these semantically equivalent forms. These relations are at the core of our ontology for the portal. The development of an ontology as a super-layer in the portal will allow us a normalization of terms used in the different businesses, without need to modify their information systems, while at the same time allowing them to effectively communicate and acquire the needed information.
Developing ontology should allow us to target specifically the ambiguity problem, while at the same time taking maximum advantage from the characteristics and benefits of the ontology [1]. The first step of the ontology development is the choice of the development methodology [10]. However, before coming to that, we need to focus and define our solution to the problem of ambiguity as understood in this work. To do so, we will define formally the equivalence relations.

5.1. Relations of Equivalences

Presently, we will no longer speak about the *terms equivalence* but of the *concepts equivalence* of the ontology. In our context, *two concepts are equivalent if they are equal in all what define them or if they are substitutable*. We will also define a *partial equivalence* which relates concepts that are not completely equivalent but are otherwise very closely related semantically in a given context. The equivalence in the literature about ontology is rather considered when talking about ontology mapping.
Definition of the equivalence relation
The notion of an equivalence relation on a set allows relating elements to be similar in a given property. Let *c1* and *c2* be two concepts in our ontology. We say that *c1* and *c2* are equivalent if there exist at least one attribute *a1* of concept *c1* and one attribute *a2* of concept *c2* such as *a1=a2*, where *a1* and *a2* are equivalent:

Equivalent (C1, C2, A) with:
C1 is a concept,
C2 is a concept different from *C1*,
A is an attribute.

We say that attribute *a1* of concept *c1* and attribute *a2* of concept *c2* are equivalent if they have the same definition and the same values. This definition, expressed formally using the descriptive logic, translates into:

$\forall$ **C1, C2, A: (C1** ***Equivalent (A)*** **C2)**
$\Leftrightarrow$ **(** $\exists$ **a1, a2 (a1** ***Attribute*** **C1) and**
(a2 ***Attribute*** **C2) and (a1=a2=A))**

« A *Attribute* C» is the relation which links a concept C to one of its attributes A.
We propose the use of ontologies and the relations of equivalence to solve the problems to bind to semantics, and precisely semantic ambiguity.
In order to build an ontology which integrates the relation of equivalence, it is necessary to use an appropriate method and to adapt it to our context.

5.2. Use and Adaptation of Methontology to Interbusiness Portals

Developing ontology for an interbusiness portal requires consideration of several characteristics using a given methodology. Among these characteristics are the information nature, its organisation within the portal, and the aimed goal through the ontology development [5].
For an information portal, it is difficult to specify accurately the type of information shared, especially if we want to preserve general applicability to this kind of portals (collaborative knowledge management

portals). At this stage of our work, we devise a method for the development of an ontology which aims at solving problems of semantic ambiguity in the exchanged terminology within an informational portal. We assume the information shared within the portal to be of the editorial type: article, dispatch, file, etc. We assume as well that the portal is built around a specific theme (field) so that to make the ontology development possible [6].

The aim of the developed ontology is to minimize problems of semantic ambiguity through the creation of semantic links, equivalence links and inequality links between various information-carrying terms. To take into account the specificity of the problem to solve, we adapt Methontology by adding an extra task at the level of conceptualization activity. This task allows the identification of the equivalence and equality relations pertaining to a given concept.

Let us consider some restrictions on the knowledge portals like leading type: article, dispatch, file, etc., and that the portal must relate to a quite precise field, so that the ontology construction will be possible.

The suggested ontology goal is to minimize the problems of semantic ambiguities while making it possible to create semantic relations, equivalence relations, and equality relations between various terms carrying information.

To take into account the specificity of the problem to be regulated, one may adapt the Methontology method:

- To orient the acquisition of knowledge to get *synonym* and *"near-synonym"*.
- To adapt the tasks of the conceptualization activity by taking account of *synonymy* and *equivalence* for the considered *concepts*.
- To add additional tasks to the conceptualization activity level, i.e. one task is considered to count the relations of equivalences and equalities relating to a given concept.

The conceptualization activities are applied as follows (the considered field is the pharmaceutical ontology which is detailed in section 6):

The conceptualization activity:

As pointed in the previous section, the conceptualization activity is the activity during which the knowledge gathered during the acquisition activity is structured. This knowledge is either present on the portal or inferred from the theme of this portal. This knowledge is structured for clarity through the different tasks of this activity.

In our context, the tasks of this activity in the collaborative knowledge portal environment are:

i) The construction of terms glossary

This glossary counts the terms carrying information to be included in our ontology, their definitions in natural language, their *synonyms* and *acronyms*. In Methontology, the terms are identified according to a Middle-out strategy (C with D which the basic terms are identified then they are specialized or generalized). In the portal, these terms can be keywords or armature of the general topic of sub-topics. Due to the portal nature between firms and the information changing flow, the activity of the knowledge acquisition takes its importance like its difficulty. In the glossary we precise (table 1):

- the terms to be included in ontology,
- their definition in natural language,
- their synonyms : complete form and shortened form; popular or commercial name for the various companies; scientific name; spelling variants; modern name and out-of-date name if it exists; terms of origin linguistic, etymological or cultural different, etc.,
- their acronyms.

Terms	Definition	Synonymes
ACIDE AMINE	Classe chimique	AMINO-ACIDE AMINOACIDE
ACIDE CAPRYLIQUE	Classe chimique	ACIDE OCTANOIQUE
ACIDE NICOTINIQUE	Classe chimique	NICOTINIQUE ACIDE
GLUCOSE	Classe chimique	DEXTROSE
ALCOOL	Classe chimique	LACTITOL
AMINOSIDE	Classe chimique	AMINOGLYCOSIDE AMINOGLYCOSIDES AMINOSIDES
BARBITURIQUE	Classe chimique	BARBITURIQUES MALONYLUREE
ENANTATE ESTER	Classe chimique	HEPTANOATE ESTER OENANTATE ESTER
LAURATE	Classe chimique	DODECANOATE
MACROLIDE	Classe chimique	MACROLIDES
ACETAMINOS-ALOL	salicylate d'acétamino-4-phényle	ACETAMINOSALOL ACETAMIDOPHENYL SALICYLATE ACETAMIDOSALOL ACETAMINOSAL CETOSALOL PHENETSAL

Table1: Sample of the term glossary for Pharmaceutical ontology.

ii) A definition of equivalence relations

This added task permits to enrich the method to allow the identification of the equivalent concepts and to count the existing types of equivalence. On the other hand, this task permits to set up a bond allowing the semantic homogenisation of our ontology concepts, and thus of the field terminology.

In this task, one will identify the relations of equivalences to be admitted for each concept.

The equality and equivalence relations are according to a certain aspect.
For example, in the pharmaceutical domain, the equivalence must keep on well defined criteria. Three types of equivalence are considered:
Therapeutic equivalence, biological equivalence, and pharmacological equivalence.
Concretely, in this task we precise (table 2):

- concept and concepts equivalent,
- type of equivalence, for example: therapeutic equivalence,
- aspect of equivalence and its values, for example: indication and Indication.

Concept	equivalents	Type of equivalence	Aspect of equivalence	Valu e
C_Chimic	C_Chimic	Eq_therapeutic	Indication	--
C_Chimic	C_Chimic	Eq_pharmacolo gic	Propriete pharmacologic	--
C_Chimic	C_Chimic	Eq_biologic	Molecul_base	--

Table2: Equivalence relations for pharmaceutical ontology.

6. A Case Study

For our case study, we are interested in pharmaceutical company's portal. We have used the portal *BIAM* [2]. We restricted our first prototype to 400 terms. The ontology is built with respect to the adapted process of Methontology.

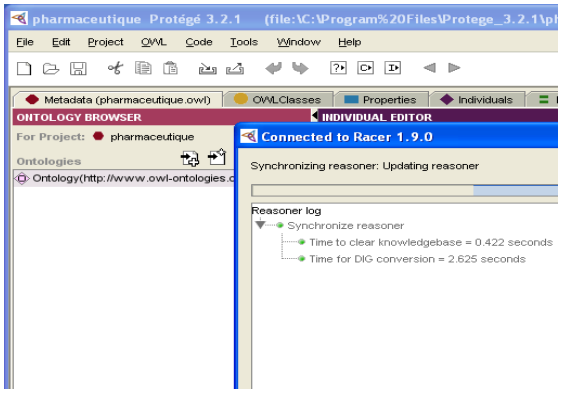

Figure 3. Implementation and validation of the ontology

For implementing this ontology (fig.3), we have used the programming language "OWL FULL" and the editor "PROTÉGÉ". Each prototype is tested conceptually by the "RACER" tool (i.e. consistency tests). The prototype obtained is an ontology being able to be enriched and supplemented. We restricted ourselves for this prototype with part of the field (the pharmacological terms). However, all terms of the field can be taken into account.

The establishment of significant criteria for the equivalence remains dependent on the study drawn up by experts of the field.
An automation of the equivalence relation will simplify the finalization of ontology.

7. Conclusion

In this work, we proposed a solution to the semantic ambiguity problem in interbusiness knowledge portals. This solution consists to build an ontology based essentially on the equivalence relations. The goal of this ontology is to be able to establish a semantic harmonization of the terminology used. We have enriched the Methontolgy method to build this ontology by adding some important tasks like the formal definition of equivalence relations. Then, we have illustrated the proposed solution on a real case study, i.e. BIAM portal. The obtained ontology prototype was implemented using OWL FULL and PROTÉGÉ, and validated using RACER.
For future work, we intend to automate the construction of the equivalence relations between the ontology concepts. Also, another perspective is to consider in the ontology construction the notion of view points which is linked to a kind of person (job, formation level, etc.) or use (the same person may be different view points according to the task that she wants to achieve.

References

[1] Barry Smith, “Ontology”. Blackwell Guide to the Philosophy of Computing and Information, Oxford: Blackwell, pp. 155-166. 2003.

[2] Banque de Données Automatisée sur les Médicaments, BIAM, www.biam2.org.

[3] Christina Tziviskou, Marco Brambilla, Politecnico di Milano, “Semantic Personalization of Web Portal Contents“. 16th International World Wide Web Conference (WWW2007) 2007.

[4] Daniel Berle, Peter Spyns, "the knowledge portal onto Web". Handbook on Ontologies, pp. 499-553. 2003.

[5] Fernandez-Lopez Mariano, "Overview of Methodologies for Building Ontologies". Proceedings of the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods. Stokholm, Sweden, August 2, 1999.

[6] Fredric Landqvist , Dick Stenmark, “Portal Information Integration and Ownership Misfits: A Case Study in a

Tourism Setting". Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06) Track8, pp. 173. 2006.

[7] Lopez, M.F.; Gomez-Perez, A.; Sierra, J.P.; Sierra, A.P., "Building a chemical ontology using Methontology and the Ontology Design Environment". Intelligent Systems and their Applications, IEEE Volume 14, Issue 1, pp. 37 – 46. 1999.

[8] Michele Missikoff, Francesco Taglino, "An Ontology-based Platform for Semantic interoperability". Handbook on Ontologies, pp. 617- 633. 2003.

[9] Rami Dib Rifaieh, "Utilisation des ontologies textuelles pour le partage sémantique entre les systèmes d'information dans l'entreprise". theses 2004.

[10] Roussey C., "Les ontologies et leurs méthodes de constructions". Séminaire au Cemagref de Clermont Ferrand. 2005.

[11] Wallace Anacleto Pinheiro, Ana Maria de C. Moura, "An Ontology Based-Approach for Semantic Search in Portals". Database and Expert Systems Applications, 15th International Workshop (DEXA'04), pp. 127-131. 2004.