Reusing Ontologies

H. Sofia Pinto and **J.P. Martins**
Instituto Superior Técnico
Departamento de Eng. Informática
Grupo de Inteligência Artificial
Av. Rovisco Pais, 1049-001 Lisboa, Portugal
Tel: (351-1) 8417641 Fax: (351-1) 8417472
sofia@gia.ist.utl.pt

Abstract

Ontology reuse is turning into an important research issue in the ontology field. Ontology reuse can be seen from two points of view: (1) assembling, extending, specializing, adapting other ontologies which are parts of the resulting ontology, or (2) merging different ontologies on the same or similar subject into a single one that unifies all of them. The first kind of reuse is named integration and is the central issue of this paper. In this article, we characterize the integration process, describe and discuss the activities that compose this process and propose an integration methodology. This integration methodology has successfully been applied to build two ontologies in different domains by reusing publicly available ontologies.

Introduction

Ontologies are used in the Knowledge Management (KM) field, for several purposes: as corporate memories, for interoperability of databases through a common ontology, etc. Therefore, results in the ontology field are of use for KM. A considerable amount of ontologies has been built during the last decade. The time has come for sharing and reusing this body of knowledge. One can see ontology reuse according to two different points of view:

- Building an ontology, by assembling, extending, specializing and adapting, other ontologies which are parts of the resulting ontology.
- Building an ontology, by merging different ontologies on the same or similar subject into a single one that unifies all of them.

Both kinds of reuse aim at building an ontology from other ontologies. Therefore, they are related to the process of ontology building. However, each kind of reuse is different from the other (Pinto, Gómez-Pérez, & Martins 1999). The first kind of reuse is named *ontology integration*. Some cases of ontology integration are described in the literature (Farquhar, Fikes, & Rice 1996; Borst, Akkermans, & Top 1997; Pinto 1999a). The second kind of reuse is named *ontology merge*. Some cases of ontology merge are also described in the literature (Wiederhold 1994; Gangemi, Pisanelli, & Steve 1998; Swartout *et al.* 1997; Noy & Musen 1999).

Our work focus on ontology integration. Although there has already been some work on the issue of ontology integration, we are still a long way from an all-purpose and foolproof integration methodology. Some of the methodologies to build ontologies (Uschold & King 1995; Gruninger 1996; Fernández *et al.* 1999) acknowledge the need for an integration step or activity but leave it undefined how integration should be performed. Some operations to help performing integration have been proposed (Farquhar, Fikes, & Rice 1996; Borst, Akkermans, & Top 1997; Pinto 1999a). However, these operations are only an initial set, since more integration operations are needed. There have also been some initial attempts to characterize integration (Pinto, Gómez-Pérez, & Martins 1999; Pinto 1999a). However, there are no ontology integration methodologies proposed in the area.

In this article we characterize integration and propose an integration methodology. In what concerns integration, we have found that rather than an activity or a step, integration is a process. We identify the activities that should be performed along this process and we characterize those activities. In what concerns the methodology, we provide guidelines and methods to perform the activities that form the integration process. These guidelines and methods, have been used in two integration experiences: building the Reference ontology (Pinto 1999a; V. *et al.* 1999; 1998) and building some of the subontologies needed to build an Environmental Pollutants ontology (Pinto 1999a; Gómez-Pérez & Rojas-Amaya 1999) reusing publicly available ontologies.

This paper addresses one important aspect of ontology reuse: ontology integration. For people involved in KM and working in corporations, ontology development is usually an intermittent activity and ontologies are just one by-product, not their main activity. Therefore, a methodology to more easily build ontologies is rather useful. The advantage of ontology integration is that, provided that a set of small, modular, highly reusable ontologies are available, large ontologies for specific purposes can more easily be assembled.[1]

We begin by introducing some terminology. Then, we briefly describe our experience in ontology integration, its main assumptions and interesting aspects. Then, we char-

[1]These small ontologies may have to be modified, adapted, prior to be assembled.

acterize the integration process, by describing and characterizing each one of the activities that form this process. We finally present an integration methodology. For each activity that composes the integration process we propose guidelines and methods to perform them.

Terminology

For us, an ontology consists of: classes, instances, relations, functions and axioms. We generically refer the union of classes and instances as *concepts*. Each one of the constituents of an ontology is generically referred as a *knowledge piece*. Each knowledge piece is associated with a name, a documentation and a definition.

Experiences in Ontology Integration

We built the Reference ontology (Pinto 1999a; V. *et al.* 1999; 1998) which is a domain ontology about ontologies (a kind of yellow pages of ontologies). This ontology gathers information, describes and has links to existing ontologies using a common organization. This ontology aims at helping future users to locate candidate ontologies for a given application. To build the Reference ontology we have reused the $(KA)^2$ ontology (Benjamins & Fensel 1998; Benjamins *et al.* 1999). $(KA)^2$ is an ontology currently under development by the Knowledge Acquisition (KA) community about itself. One of the research topics and products of the KA community is ontologies. Therefore, they were already represented in $(KA)^2$. However, knowledge about ontologies in $(KA)^2$ was represented so as to allow one to distinguish ontologies from other KA research topics. The Reference ontology, which reused $(KA)^2$, introduced knowledge that allowed one ontology to be distinguished from another. Therefore, this knowledge was introduced for ontologies described as products. One interesting aspect of this experience was the fact that some of the features that are important to characterize ontologies were also important to describe products in general. Therefore, some of these features were relocated to characterize products in general and not only ontologies in particular. Another interesting feature of this experience is the fact that ontologists acted both as domain experts and as ontologists since the domain of the Reference ontology is ontologies, their domain of expertise.

We were involved within a vast interdisciplinary team effort to build an Environmental Pollutants ontology (Pinto 1999a; Gómez-Pérez & Rojas-Amaya 1999). This ontology aims at representing a unified complete and consistent set of concepts needed to build environmental applications. The Environmental Pollutants ontology (EPO) is an ongoing project, however its most important subontologies have already been built. One of these is the Monoatomic Ions ontology. This ontology represents knowledge about monoatomic ions from a general, a chemical and an environmental point of view. To build the Monoatomic Ions ontology both Chemicals (Fernández *et al.* 1999) and Standard-Units (Gruber & Olsen 1994) ontologies were reused. One of the steps in the development of this ontology was the evaluation for integration purposes of the ontologies to be reused by domain experts. In the case of Chemicals, the chemical and environmental experts recommended a series of changes to be performed. These were carried out by the ontologists.[2] For instance, one of the conclusions of the domain experts was that only Chemical-Elements, the subontology of Chemicals that defines the chemical elements of the periodic table, should be reused. All reused ontologies were verified, revised and corrected according to the domain experts suggestions. In the case of Chemical-Elements, its revised and corrected version was included in the Monoatomic Ions ontology.

An important point to be stressed out from all of our experiences, is that we had access to the knowledge level representations of all reused ontologies (for instance, $(KA)^2$, Chemicals, etc.). We would like to point out that all of our work was done at the knowledge level. This simplified the overall process of integration. If the knowledge level representation of an ontology is not available, then an ontological reengineering process (Blázquez *et al.* 1998) can be applied.

We would also like to point out that in both cases there was no need to translate ontologies between different knowledge representation languages. The translation of ontologies is in itself a very important and difficult problem to be solved in order to allow more generalized reuse of ontologies. As discussed in (Uschold *et al.* 1998), translation is far from being a fully automatic process in the near future.

The Integration Process

In this section we present the most important conclusions about integration. The main conclusion is that rather than a step or an activity, as previous ontology building methodologies (Uschold & King 1995; Gruninger 1996; Fernández *et al.* 1999) had proposed, integration is a process that takes place along the entire ontology building life cycle. This process involves one or more ontologies that are integrated and a resulting ontology. In an integration process one can identify in the resulting ontology regions that were taken from the integrated ontologies. Knowledge in those regions was left more or less unchanged.

As any process, integration is composed of several activities. We have identified the activities that take place along the ontology building life cycle to perform integration. We would like to point out that the development of an ontology follows an evolving prototyping life cycle. Therefore, the ontology may be considered for integration in specification, conceptualization, formalization, implementation and maintenance steps.[3] For this reason one can have integration activities for the same ontology in different stages of the ontology building process.

Another important conclusion is that integration should begin as early as possible in the ontology building life cycle so that the overall ontology building process is simplified. In both our cases, integration began as early as the conceptual-

[2] See (Pinto 1999a) for a description of the changes that were recommended by the experts.

[3] These are the ontology building steps usually acknowledged by the most important ontology building methodologies. Activities that may take place along the whole ontology building life cycle include knowledge acquisition and evaluation.

ization phase. One of the consequences of this conclusion is that more integration effort should be made at the earliest stages (specification and conceptualization) than in final ones (implementation or maintenance). Usually only implemented ontologies are made available at ontology libraries. If one begins integration as early as conceptualization, then one needs ontologies represented at the knowledge level, not at the implementation level. Therefore, an ontological reengineering process (Blázquez *et al.* 1998) may have to be applied. This process is usually composed of three steps: a reverse engineering step, a restructuring step and a forward engineering step. To get the knowledge level representation of an implemented ontology either the first or the two initial steps are applied. For instance, while building the EPO ontology one of the reused ontologies, the EngMath ontology (Gruber & Olsen 1994), was only available at the implementation level. Therefore, an ontological reengineering process (Gómez-Pérez & Rojas-Amaya 1999) was conducted.

As we said, we identified the several integration activities that take place during the ontology building process. The initial activities to perform integration include: finding and choosing candidate ontologies to be reused, evaluation of the candidate ontologies by domain experts through specialized criteria oriented to integration, assessment of the candidate ontologies by ontologists through specialized criteria oriented to integration, and choosing the most adequate ontology to be reused among the analyzed candidate ontologies. These activities precede integration of knowledge from the integrated ontology into the resulting ontology. They help the ontologist to analyze, compare, and choose the ontologies that are going to be reused. The specialized criteria used in integration oriented evaluation and assessment enhance the possible problems that a particular ontology may have in a particular integration process. They allow ontologists and domain experts to identify and be aware of those problems. The ontology chosen to be reused may lack knowledge, may require that some knowledge is removed, etc., that is, it may not exactly be what is needed. The best candidate ontology is the one that can best (more closely) or more easily (using less operations) be adapted to become the needed ontology.

When this process ends, that is the ontology appropriate to be reused for this particular integration process is found, we must integrate the knowledge of that ontology. For that, one needs integration operations. These can be viewed as composing, combining, modification or assembly operations. Integration operations specify how knowledge from an integrated ontology is going to be included and combined with knowledge in the resulting ontology, or modified before its inclusion. Knowledge from integrated ontologies can be, among other things, (1) used as it is, (2) adapted (or modified), (3) specialized (leading to a more specific ontology on the same domain) or (4) augmented (either by more general knowledge or by knowledge at the same level). We have identified the basic integration operations that are needed to perform integration (Pinto 1999a; 1999b). These are the basis for the more complex integration operations (Pinto 1999b). All identified operations are organized in a taxonomy of integration operations.

After integration of knowledge, one should evaluate, assess and analyze the resulting ontology. Besides the usual criteria involved in evaluation and assessment of an ontology (Gómez-Pérez, Juristo, & Pazos 1995), one should pay attention to a specialized set of criteria that specifically analyzes whether the resulting ontology has enough quality.

One important aspect of integration is the fact that this process is included in the overall ontology building process. If an ontology adequate to be reused is not found, then one must build it from scratch using one of the available ontology building methodologies. In general, we are looking for and choosing the subontologies that compose the final ontology since, in general, an ontology is composed of several "modules". If we find an ontology that matches the whole ontology that one needs to build, then one does not need to apply integration operations or analyze the resulting ontology. However, finding candidate ontologies, their evaluation and assessment for integration purposes, and the choice of the most adequate one remain essential activities to be performed. It should be noted that only the most important integration activities were referred. Other activities, some of which are also important in any ontology building methodology, include: identification of the modules in which the ontology can be divided into, identification of the assumptions and ontological commitments that those modules should comply to, etc. For the first abstraction capabilities are used. For the second the specification requirements document provides an adequate source of information. In integration the modules are obviously related to candidate ontologies.

As a summary, the most important activities to be performed in an integration process include:

- finding and choosing candidate ontologies to be integrated,
- evaluation of the candidate ontologies by domain experts through specialized integration criteria,
- assessment of the candidate ontologies by ontologists through specialized integration criteria,
- choosing the adequate ontology to be integrated,
- application of the appropriate integration operations,
- analysis of the resulting ontology.

An Integration Methodology

In the following subsections we describe procedures, guidelines and methods to help perform each one of the activities that compose the integration process.

Finding and Choosing Candidate Ontologies

Some of the activities that form the integration process are looking for, finding and choosing candidate ontologies to be reused. In both our experiences, not only had we previous knowledge of the reused ontologies from articles in the literature, but they were also available at the Ontolingua Server library (Farquhar, Fikes, & Rice 1996). Therefore, we had access to study whether they were appropriate to be reused and to reuse them. To help at finding ontologies to be reused, we were involved in the construction of $(\text{ONTO})^2$Agent (V.

et al. 1998), a WWW broker. This Web Agent[4] keeps a characterization of ontologies, according to a pre-defined taxonomy of features (V. *et al.* 1998) (from which the Reference ontology was built). The characterization of ontologies according to those features is provided by their developers. To look for ontologies, the user asks the agent to perform a search and the ontologies whose characterization matches the characterization of that search are displayed to the user.

The ontologies to be analyzed, and perhaps reused, are chosen from those available in libraries that meet a series of requirements, such as, it is about the adequate domain, it is of the adequate type (van Heist, Schreiber, & Wielinga 1997), it was evaluated, etc. The ontologies chosen as candidates must also be compatible with the requirements specified for the resulting ontology. In the case of domain, the candidate ontology must be of the appropriate domain. For instance, if we need an ontology about mathematics, we have no use for an ontology about cars. However, not all requirements need a perfect match. For instance, in what regards evaluation, if the ontology has been evaluated, then it probably has more quality than other ontologies about the same domain that were not evaluated. This does not mean that an unevaluated ontology cannot be reused. In fact, in an integration process some of its activities involve the complete analysis (evaluation and assessment) of the ontology from an integration point of view. Therefore, the ontology quality problems will be identified later.

In conclusion, we have strict (hard, "must have") requirements, such as domain, and desirable (soft, "good to have") requirements, such as evaluated. Strict requirements eliminate ontologies that are not appropriate to be reused. Desirable requirements distinguish between those ontologies that are more likely to be better candidates, but they do not exclude candidate ontologies.

Integration Oriented Evaluation

Another activity of the integration process is evaluation of the candidate ontologies from an integration point of view. We have identified a series of criteria that domain experts should take into account when analyzing an ontology for integration. Domain experts should evaluate the ontology paying special attention to:

what knowledge is missing (by knowledge we mean any knowledge piece, such as, concepts, classification criteria, relations, etc.). Sometimes some knowledge pieces that are relevant, important and usually used to characterize the domain are not represented in the ontology. This includes not only classes and instances, but also important distinctions made of the domain concepts, different classification criteria that are widely accepted to characterize the domain, relations that are relevant to represent knowledge about the domain (which relations should be specified and for which concepts should they be specified), etc. The domain experts should also analyze what important knowledge about the domain is missing in the ontology in view to the particular use that the ontology is going to have. For instance, while we were building the Reference ontology we have found that some concepts that are important to describe ontologies in general were missing in KA^2, such as the class of `ontology-languages` or one of its instances, the language `ontolingua`.

what knowledge should be removed Sometimes some knowledge pieces represented in the ontology are superfluous, either because they are not important, or not relevant, or not usually used to describe the domain in question or because they are not needed for the particular use that the ontology is going to have. For instance, while we were building the Monoatomic Ions ontology the chemical and environmental experts that evaluated Chemicals concluded that only one of its subontologies, `Chemical-Elements`, should be reused because the crystalline structure of the chemical elements (the other subontology of Chemicals) is not relevant from an environmental point of view.

what knowledge should be relocated Sometimes knowledge pieces should be placed elsewhere in the ontology so that the domain is best characterized. For instance, in the case of the Reference ontology, some of the properties and relations that were found important to characterize ontologies were also important to characterize products in general. Therefore, these properties and relations were relocated. For instance, the relation `distributed-by` found important to characterize an ontology was established between a `product` and an `organization` (representing its distributor) and not specifically between an `ontology` and an `organization`.

which knowledge sources should be changed Sometimes some of the knowledge sources used to acquire knowledge are not the most reputable or up-to-date. Knowledge from those sources that is represented in the ontology should be replaced by more reputable, standard and more up-to-date knowledge. For instance, in the case of Chemical-Elements the chemical experts questioned the knowledge sources used for the values of the chemical properties of the chemical elements. The values of some of those properties were not from the most reputable sources or were outdated or were measured in less standard (usual) scales. In this case the experts suggested that the values of the properties from (Barbor & Ibarz 1979) should be changed for the values in (Greenwood & Earnsshaw 1986). For instance, the values of the property `Density-at-20°C` were changed accordingly.

which documentation should be changed Sometimes the documentation of the domain is not correct (syntactically and semantically), precise, complete (comprehensive) or reflects the last discoveries in the field and should be changed. The documentation should explain the domain and the knowledge pieces represented in the ontology so that a non-expert could learn enough about the domain to be able to understand the concepts that are represented in the ontology. For instance, in the case of Chemical-Elements the chemical experts questioned the documenta-

[4] $(ONTO)^2$Agent is composed of OntoAgent (a WWW broker) and the Reference Ontology. $(ONTO)^2$Agent is available at `http://delicias.dia.fi.upm.es/OntoAgent/index.html`.

tion of some of the concepts represented in the ontology, specially the documentation of the concepts represented in its upper-levels (representing the classification criteria used to structure Chemical-Elements). For instance, the documentation of the definition of the class `actinides` was replaced by the definition presented in (Greenwood & Earnsshaw 1986).

which terminology should be changed Sometimes the terminology used is not the most usually accepted in a (sub)field or in a related field, or the standard terminology and should be changed. For instance, in the case of Chemical-Elements the chemical experts questioned some of the terminology that was used. Not only was some of the terminology not the most usual or adequate one but also some did not comply to the standard terminology recommended by the International Union of Pure and Applied Chemistry (IUPAC). For instance, the property `Resistivity-At-20-C` should be renamed `Electrical-Resistivity-At-20-C` and the chemical element `Wolfram` should be renamed `Tungsten` as specified by IUPAC.

which definitions should be changed Sometimes the definitions used are not the most usually accepted, standard or composed of the definitional characteristics of the knowledge pieces. For instance, in the case of KA^2 the class `Person` was inappropriately defined with the binary relation `Member-of-Program-Committee`. This relation should not be used in the definition of this concept since this is not a typical attribute of this concept. It should only be used in the definitions of those concepts where it is a typical attribute.

which practices should be changed Sometimes the procedures used to gather knowledge (knowledge acquisition) and to build the ontology (ontological engineering) are not the most correct ones or follow the accepted best practices in the domain area. For instance, in the case of Chemical-Elements all values for each chemical property should be taken from one single knowledge source for all the elements. Therefore, if no value is provided in the adopted knowledge source for one particular element, then it should not be looked for in another source. Since this procedure had not been followed while building Chemical-Elements, the ontology had to be revised and changed to assure the enforcement of this best practice.

These evaluation criteria are different from those through which an ontology should be evaluated during its ontology building life cycle. They are oriented towards the integration of knowledge that will be performed in subsequent stages of this process. They show the weaknesses and strong points of the candidate ontology from the domain expert point of view.

Integration Oriented Assessment

Another activity of the integration process is assessment of the candidate ontologies from an integration point of view. We have identified a series of criteria that ontologists should take into account when analyzing an ontology for integration. Ontologists should assess the ontology paying special attention to:

general structure of the ontology It is important to assess whether the general structure of the ontology is appropriate, that is, it complies with the requirements need of it. To analyze the structure of the candidate ontology six criteria should be taken into consideration:

- Is the structure adequate (one hierarchy, several hierarchies, a graph, etc.) and preferably well-balanced?
- Is the ontology divided into adequate and enough modules, that is, is the ontology divided into natural and appropriate (quality and quantity) subontologies?
- Is there adequate and enough specialization of concepts, that is, are the needed concepts and their specializations represented?
- Is inheritance correctly and appropriately used?
- Is there enough diversity represented in the ontology so that new concepts are more easily introduced?
- Are similar concepts represented close to one another whereas less similar concepts are represented further apart (minimization of the semantic distance between sibling concepts (V. *et al.* 1998))?

For instance, while integrating EngMath (Gruber & Olsen 1994) to build the EPO ontology, part of its structure was found ill-balanced (all instances of Standard-Units were subclasses of `unit-of-measure`) (Gómez-Pérez & Rojas-Amaya 1999) and poor (not enough diversity), although adequate and divided into appropriate modules. In $(KA)^2$ inheritance was being inappropriately used (Pinto 1999a). If the ontology has not the adequate structure, then the changes to be made can be so extensive that it may be more cost effective to build an ontology from scratch. This is why its analysis is so important.

basic distinctions Are the relevant and required (quantity and quality) basic distinctions (classification criteria made of the concepts described in the ontology) represented? Changing the basic distinctions (usually represented at the top-levels of the ontology) upon which the ontology is based can also imply a vast revision of the ontology. For instance, changing the top-level of Chemical-Elements according to natural state at normal pressure and temperature conditions (solid, liquid, gaseous) instead of the group classification proposed by Mendeleev (IA, IIA, etc.) will cause profound changes in the whole ontology. The only knowledge that can be reused are the instances representing the elements. Usually, in this case, it is preferable to build another ontology.

structuring relation Is the privileged relation upon which the ontology is structured the required one[5]? Changing the privileged relation according to which the ontology is organized can also have profound consequences. The whole knowledge would, most probably, have to be revised, since the new relation organizes knowledge in a

[5]An ontology can be thought of as structured or organized according to one privileged relation, for example, ISA, part-of, etc.

completely different way. Knowledge about a given domain that should be represented using one relation has nothing to do with what should be represented using another relation. Probably it is preferable to build a new ontology from scratch (if none of the available ones meets our needs). For instance, changing the $(KA)^2$ ontology, which is structured upon the `ISA` relation, so that it is structured upon the `part-of` relation entails building a new ontology about the KA community.

naming convention rules Do names of knowledge pieces follow standardization rules? For instance, in the Reference ontology whenever possible binary relations were named by concatenating the name of the ontology (or the name of the concept representing the first element of the relation), the name of the relation and the name of the target concept. For instance, the relation `ontology-formalized-in-language` between an `ontology` and an `ontology-language`. However, relations in $(KA)^2$ did not follow such rules. Whenever possible, naming convention rules should be enforced all over the resulting ontology so that terminology becomes coherent. This increases reusability and usability of the resulting ontology (it is easier to find relevant knowledge and it is easier to introduce new knowledge).

definitions Do the definitions of the knowledge pieces follow unified patterns, are clear, concise, consistent, complete, correct (lexically and syntactically), precise and accurate? Are they efficient? All these questions deal with the way knowledge is represented in the ontology. For instance, in the case of $(KA)^2$ there were some minor errors (cut and paste errors) in the definitions of some concepts (that were corrected). Although minor errors, they lead to incorrect definitions. For instance, in the case of Standard-Units some definitions did not follow unified patterns (Gómez-Pérez & Rojas-Amaya 1999).

documentation Is the documentation clear, helpful and adequate? Does it discuss alternative representations and the choices that were made to represent knowledge? Is it coherent in relation to the definition of the knowledge piece? Although documentation is one of the constituents of an ontology knowledge piece, it usually is its most neglected component. For instance, in the $(KA)^2$ version at the Ontolingua Server,[6] the documentation is not associated to the knowledge pieces, although it is available in the textual version of the Research-Topic subontology.[7]

knowledge pieces represented Are all and only the appropriate knowledge pieces represented (or included)? This issue should be analyzed taking into account the knowledge pieces that domain experts have found important to represent. If they have found some knowledge pieces lacking/superfluous they should be added/deleted to/from the ontology. The use made of the knowledge in the ontology also influences the way those knowledge pieces are represented and which knowledge pieces need to be represented. The ontologist analysis has to focus both aspects: the relevant and needed knowledge pieces are represented and they are usefully represented.

These assessment criteria are different from those through which an ontology should be assessed during its ontology building life cycle. They are oriented towards the integration of knowledge that will be performed in subsequent stages of this process. They show the weaknesses and strong points of the candidate ontology from the knowledge representation point of view.

Choosing the Adequate Ontology

After the analysis of the various candidate ontologies, and given the fact that they may not perfectly match the needed ontology, another choice must take place. Among the candidate ontologies that passed strict requirements and among those that scored best in integration oriented evaluation and assessment, one has to choose the ontology that best suits our needs, or that can more easily or better be adapted to them. This choice also depends to some extent on the other ontologies that are going to be reused since in an integration process one can reuse more than one ontology. It is important that the reused ontologies are compatible among themselves, namely in what concerns their assumptions, for instance, their ontological commitments (Gruber 1995).

Integration Operations

To integrate knowledge one needs integration operations. We have identified and defined a taxonomy of integration operations at the knowledge level (Pinto 1999b; 1999a). Ontology integration operations are classified into basic and non-basic (complex) operations. An algebra of basic integration operations was defined (Pinto 1999b). Therefore, symbolic or implementational issues were not taken into consideration. Basic integration operations are divided into:

operations on whole ontologies such as, inclusion of an ontology.

operations on the constituents of an ontology are subdivided into operations that:

- **remove or introduce a knowledge piece** so that the ontology contains the appropriate and needed knowledge to describe the domain,
- **change the name of a knowledge piece** to comply to naming convention rules, or introduce standard or more usual terminology,
- **change the documentation of a knowledge piece** to update, correct or increase its clarity,
- **change the definition of a knowledge piece** to more accurately, simply, clearly, correctly, precisely, completely, etc. represent knowledge of a given domain.

Changes refer to removal, introduction or composition of removal and introduction operations. We have operations that introduce or remove a class, instance, relation, etc., its name, its documentation and its definition. All these operations should be used parsimoniously otherwise it may

[6] At Madrid's mirror site, `http://www-ksl-svc-lia.dia.fi.upm.es:5915` account `''ontologias-ka2''`.

[7] `http://www.aifb.uni-karlsruhe.de/WBS/broker/KA2.html`.

be more cost-effective to build the ontology from scratch. Special care must be exerted with the operations that relate knowledge pieces to other knowledge pieces. For instance, inclusion of a class usually needs the specification of the place where the concept should be introduced (for instance, by stating the parents of the class) and removal of a class may imply or not the removal of the whole hierarchy underneath it. Changes in the definition include, a series of other changes. For instance, in the case of concepts (classes or instances), it includes changes in the relations in which the concept is involved (including the privileged relations that structure the ontology), changes in the properties defining the concept, etc. These basic integration operations are the basis for the non-basic integration operations such as ontology mapping (Borst, Akkermans, & Top 1997), relocation (Pinto 1999b), subontologize (Pinto 1999b), collapsing (Pinto 1999b), etc.

Resulting Ontology

The resulting ontology should meet a series of requirements, besides the usual evaluation (verification and validation) and assessment criteria (Gómez-Pérez, Juristo, & Pazos 1995) (which include completeness, conciseness, consistency, expandability and robustness) and the usual features of a good ontology (Gruber 1995) (clarity, coherence, extendibility, minimal encoding bias, minimal ontological commitment). The resulting ontology should be, among other features, non-ambiguous (related to coherence), be built upon the appropriate basic distinctions, be consistent and coherent all over (although composed by knowledge from different integrated ontologies), and have both an adequate (enough) level of detail and an appropriate level of detail throughout the whole ontology, that is, there are no islands of exaggerated level of detail and other parts with an adequate one.[8]

Conclusions and Future Work

In this article we present the characterization of the ontology integration process and propose a methodology to carry it out. The important activities composing the methodology are described. They include: finding and choosing candidate ontologies, integration oriented evaluation and assessment of candidate ontologies, choosing adequate ontologies to be integrated, application of integration operations to integrate knowledge and analysis of the resulting ontology. Guidelines and methods to perform these activities are given.

One of the advantages of the proposed integration methodology is its generality which makes possible the use of this methodology with different ontology building methodologies. The only assumption made by this methodology is that knowledge should be represented at the knowledge level. This methodology has successfully been applied to build two ontologies in two different domains by means of integration. In both cases, the integration methodology was used in conjunction with METHONTOLOGY framework (Fernández *et al.* 1999), which is an ontology building methodology that enables the construction of ontologies at the knowledge level. In the future, we would like to use the proposed methodology in conjunction with other ontology building methodologies, such as (Uschold & King 1995), to build ontologies by means of integration. We would also like to use this methodology to build and reuse Knowledge Management ontologies.

[8]It should be stressed that none of the parts should have less level of detail than the required one or else the ontology would be useless, since it would not have sufficient knowledge represented.

Acknowledgments

We thank all members of GIA at Instituto Superior Técnico for their support, specially António Leitão. Both the Reference Ontology and the Environmental Pollutants Ontology were built during our stay at the Knowledge Sharing and Reuse group at the Lab. de Inteligencia Artificial of the Universidad Politecnica de Madrid (UPM). We thank Prof. Asunción Gómez-Pérez for the opportunity to work within her group.

References

Barbor, J. A., and Ibarz, J. 1979. *Quimica General Moderna.* Martin.

Benjamins, R., and Fensel, D. 1998. The Ontological Engineering Initiative $(KA)^2$. In Guarino, N., ed., *Formal Ontology in Information Systems*, 287–301. IOS Press.

Benjamins, R.; Fensel, D.; Decker, S.; and Gómez-Pérez, A. 1999. $(KA)^2$: Building Ontologies for the Internet, a Mid Term Report. *International Journal of Human Computer Studies* 51:687–712.

Blázquez, M.; Fernández, M.; García-Pinar, J. M.; and Gómez-Pérez, A. 1998. Building Ontologies at the Knowledge Level Using the Ontology Design Envirnoment. In *Knowledge Acquisition Workshop.*

Borst, P.; Akkermans, H.; and Top, J. 1997. Engineering Ontologies. *International Journal of Human Computer Studies special issue "Using Explicit Ontologies in KBS Development"* 46(2/3):365–406.

Farquhar, A.; Fikes, R.; and Rice, J. 1996. The Ontolingua Server: A Tool for Collaborative Ontology Construction. In *Knowledge Acquisition Workshop.*

Fernández, M.; Gómez-Pérez, A.; Sierra, A. P.; and Sierra, J. P. 1999. Building a Chemical Ontology Using METHONTOLOGY and the Ontology Design Environment. *IEEE Intelligent Systems* 14(1):37–46.

Gangemi, A.; Pisanelli, D. M.; and Steve, G. 1998. Ontology Integration: Experiences with Medical Terminologies. In Guarino, N., ed., *Formal Ontology in Information Systems*, 163–178. IOS Press.

Gómez-Pérez, A., and Rojas-Amaya, D. 1999. Ontological Reengineering for Reuse. In *11th European Workshop on Knowledge Acquisition, Modeling and Management, EKAW99.* Springer Verlag.

Gómez-Pérez, A.; Juristo, N.; and Pazos, J. 1995. Evaluation and Assessement of the Knowledge Sharing Technology. In Mars, N., ed., *Towards Very Large Knowledge Bases*, 289–296. IOS Press.

Greenwood, N. N., and Earnsshaw, A. 1986. *Chemistry of the Elements*. Pergamon Press.

Gruber, T., and Olsen, G. R. 1994. An Ontology for Engineering Mathematics. In J. Doyle, E. S., and Torasso, P., eds., *KR94 Proceedings*, 258–269. Morgan Kaufmann.

Gruber, T. 1995. Towards Principles for the Design of Ontologies for Knowledge Sharing. *International Journal of Human Computer Studies* 43(5/6):907–928.

Gruninger, M. 1996. Designing and Evaluating Generic Ontologies. In *ECAI96's workshop on Ontological Engineering*.

Noy, N. F., and Musen, M. A. 1999. An Algorithm for Merging and Aligning Ontologies: Automation and Tool Support. In *AAAI99's workshop on Ontology Management, WS-99-13*, 17–27. AAAI Press.

Pinto, H. S.; Gómez-Pérez, A.; and Martins, J. P. 1999. Some Issues on Ontology Integration. In *IJCAI99's workshop on Ontologies and Problem Solving Methods: Lessons Learned and Future Trends*.

Pinto, H. S. 1999a. Towards Ontology Reuse. In *AAAI99's workshop on Ontology Management, WS-99-13*, 67–73. AAAI Press.

Pinto, H. S. 1999b. Towards operations to ontology integration. Technical Report GIA 99/02, Grupo de Inteligência Artificial do Instuto Superior Técnico.

Swartout, B.; Patil, R.; Knight, K.; and Russ, T. 1997. Toward Distributed Use of Large-Scale Ontologies. In *AAAI97 Spring Symposium Series, workshop on Ontological Engineering*.

Uschold, M., and King, M. 1995. Towards a Methodology for Building Ontologies. In *IJCAI95's workshop on Basic Ontological Issues in Knowledge Sharing*.

Uschold, M.; Healy, M.; Williamson, K.; Clark, P.; and Woods, S. 1998. Ontology Reuse and Application. In Guarino, N., ed., *Formal Ontology in Information Systems*, 179–192. IOS Press.

V., J. A.; Gomez-Perez, A.; T., A. L.; and Pinto, H. S. 1998. (ONTO)^{2}Agent: An Ontology-Based WWW Broker to Select Ontologies. In *ECAI98's workshop on Application of Ontologies and Problem Solving Methods*.

V., J. A.; Gomez-Perez, A.; T., A. L.; and Pinto, H. S. 1999. How to Find Suitable Ontologies Using an Ontology-Based WWW Broker. In Mira, J., and Sanchéz-Andrés, J., eds., *Engineering Applications of Bio-Inspired Artificial Neural Networks, Lecture Notes in Computer Science, number 1607*, 725–739. Springer Verlag.

van Heist, G.; Schreiber, A. T.; and Wielinga, B. J. 1997. Using Explicit Ontologies in KBS Development. *Internacional Journal of Human-Computer Studies* 45:183–292.

Wiederhold, G. 1994. Interoperation, Mediation and Ontologies. In *Proceedings of the International Symposium on the Fifth Generation Computer Systems, Workshop on Heterogeneous Cooperative Knowledge-Bases*, volume W3, 33–48.