Ontology Integration: How to perform the Process*

Helena Sofia Pinto and **João P. Martins**
Grupo de Inteligência Artificial
Departamento de Eng. Informática
Instituto Superior Técnico
Universidade Técnica de Lisboa
Av. Rovisco Pais
1049-001 Lisboa, Portugal

Abstract

Although ontology reuse is an important research issue only one of its subprocesses (merge) is fairly well understood. The time has come to change the current state of affairs with the other reuse subprocess: integration. In this paper we characterize the ontology integration process, we identify the activities that should be performed in this process and describe a methodology to perform the ontology integration process.

1 Introduction and motivation

Ontologies aim at capturing static domain knowledge in a generic way and provide a commonly agreed upon understanding of that domain, which may be reused and shared across applications and groups [Chandrasekaran *et al.*, 1999]. Therefore, one can define an ontology as a shared specification of a conceptualization. Ontology reuse is now one of the important research issues in the ontology field. There are two different reuse processes [Pinto *et al.*, 1999]: (1) *merge* and (2) *integration*. Merge is the process of building an ontology in one subject reusing two or more different ontologies on that subject [Pinto *et al.*, 1999]. In a merge process the source ontologies are unified into a single one, so it usually is difficult to identify regions in the resulting ontology that were taken from the merged ontologies and that were left more or less unchanged.[1] It should be stressed that in a merge process the source ontologies are truly different ontologies and not simple revisions, improvements or variations of the same ontology. Integration is the process of building an ontology in one subject reusing one or more ontologies in different subjects[2] [Pinto *et al.*, 1999]. In an integration process source ontologies are aggregated, combined, assembled together, to form the resulting ontology, possibly after reused ontologies have suffered some changes, such as, extension, specialization or adaptation. 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.

A lot of research work has been conducted under the merge area. There is a clear definition of the merge process [Sowa, 2000], operations to perform merge have been proposed [Noy and Musen, 1999; Wiederhold, 1994], a methodology is available [Gangemi *et al.*, 1998] and several ontologies have been built by merging several ontologies into a single one that unifies all of the reused ontologies [Swartout *et al.*, 1997; Gangemi *et al.*, 1998]. The first tools to help in the merge process are now available [Noy and Musen, 2000; McGuiness *et al.*, 2000].

In the integration area a similar effort is now beginning. The most representative ontology building methodologies [Uschold and King, 1995; Gruninger, 1996; Fernández *et al.*, 1999] recognize integration as part of the ontology development process, but none really addresses integration. Integration is only recognized as a difficult problem to be solved. They don't even agree on what integration is: for some it is an activity, for others it is a step. We have been involved in two integration experiences where publicly available ontologies were reused: we built the Reference ontology [Arpirez-Vega *et al.*, 2000; Pinto and Martins, 2000; Pinto, 1999a; Arpirez-Vega *et al.*, 1998] and we were involved in building some of the subontologies needed to build an Environmental Pollutants ontology (EPO) [Pinto and Martins, 2000; Pinto, 1999a; Amaya, 1998; Gómez-Pérez and Rojas-Amaya, 1999].

We have found that integration is far more complex than previously hinted. It is a process of its own [Pinto, 1999a; Pinto and Martins, 2000]. In this article we characterize integration, we identify the activities that should be performed in this process and we characterize those activities. We describe the methodology that we developed to perform the activities that form this process.

2 Terminology and assumptions

Ontology building is a process that follows an evolving prototyping life cycle. The usually accepted *stages* through which

* This work was partially supported by JNICT grant No. PRAXIS XXI/BD/11202/97 (Sub-Programa Ciência e Tecnologia do Segundo Quadro Comunitário de Apoio).

[1] In some cases, knowledge from the merged ontologies is homogenized and altered through the influence of one source ontology on another (is spite of the fact that the source ontologies do influence the knowledge represented in the resulting ontology). In other cases, the knowledge from one particular source ontology is scattered and mingled with the knowledge that comes from the other sources.

[2] The subjects of the different ontologies may be related.

an ontology is built are:[3] specification, conceptualization, formalization, implementation, and maintenance. At each stage there are *activities* to be performed. Besides the activities of *specification*, in which one identifies the purpose (why is the ontology being built?) and scope (what are its intended uses and end-users?) of the ontology, *conceptualization*, in which one describes, at a conceptual level, the ontology that should be built so that it meets the specification found in the previous step, *formalization*, in which one transforms the conceptual description into a formal model, *implementation* in which one implements the formalized ontology in a formal knowledge representation language, and *maintenance*, in which one updates and corrects the implemented ontology; that should be performed at each homonymous stage, there are other activities, such as, *knowledge acquisition*, in which one acquires knowledge about the domain either by using elicitation techniques on domain experts or by referring to relevant bibliography, *documentation*, in which one reports in a document and along the implementation, what was done, how it was done and why it was done, *integration*, in which one reuses other ontologies as much as possible, and *evaluation*, in which one technically judges the ontology.

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

The aim of the conceptualization phase is to describe in a conceptual model the ontology that should be built. We assume that, in this phase of *any* ontology building process questions like,

- what should be represented in the ontology?
- how should it be represented (as a class, relation, etc.)?
- which relation should be used to structure knowledge in the ontology?
- which structure is the ontology going to have (graph, tree, etc.)?
- which ontological commitments and assumptions should the ontology comply to?
- which knowledge representation ontology should be used?
- should the ontology be divided in modules?
- in which modules should the ontology be divided?

are answered.

3 The Process

In this section we present the most important conclusions about integration and its characterization.

3.1 Main findings

The main conclusion is that integration is a process that takes place along the entire ontology building life cycle, rather than a step or an activity, as previous ontology building methodologies proposed [Pinto, 1999a; Pinto and Martins, 2000].

As any process, integration is composed of several activities. We have identified the activities that should take place along the ontology building life cycle to perform integration. Since the development of an ontology follows an evolving prototyping life cycle, integration activities can take place for one ontology in any stage 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 [Pinto, 1999a; Pinto and Martins, 2000]. In both our cases, integration began as early as the conceptualization phase. Since in conceptualization much of the design of the ontology is specified, it is considerably more difficult to try to integrate an ontology at the implementation phase because, unless one has prior knowledge of the ontologies available for reuse, available ontologies will rarely match the needs and the conceptual model found for the resulting ontology. One of the consequences of this conclusion is that more integration effort should be made at the earliest stages, specially in conceptualization and formalization, than at final ones, implementation or maintenance [Pinto and Martins, 2000].

At the conceptualization phase, one uses knowledge level [Newell, 1982] representations of ontologies. Usually, the knowledge level representation of an ontology is not publicly available (only implemented ontologies are available at ontology libraries). If the knowledge level representation of an ontology is not available, then an ontological reengineering process [Blázquez *et al.*, 1998] can be applied to get the conceptual model of an implemented ontology. This process returns one possible[4] conceptual model of an implemented ontology. When one begins integration as early as conceptualization, one needs the ontologies that are going to be considered for integration represented in an adequate form. Any conceptual model representation is adequate. An important point to be stressed out from all of our experiences is the fact that we had access to knowledge level representations of most reused ontologies as proposed by METHONTOLOGY [Fernández *et al.*, 1997]. In the case of (KA)^2 [Benjamins and Fensel, 1998; Benjamins *et al.*, 1999] (to build the Reference ontology) and Chemicals [Gómez-Pérez *et al.*, 1996; Fernández *et al.*, 1999] (to build the Monoatomic Ions subontology of EPO) we had access to the actual conceptual models that produced their Ontolingua versions, but, in the case of EPO a reengineering process was applied [Gómez-Pérez and Rojas-Amaya, 1999] to produce one conceptual model of Standard Units [Gruber and Olsen, 1994]. However, any knowledge level representation would be appropriate. Moreover, due to the particular framework that was used, ODE [Fernández *et al.*, 1999], all

[3] We use the terminology proposed in [Fernández *et al.*, 1999] since it is the most consensual in the field.

[4] It should be stressed that this process may not produce the actual conceptual model that originated the final ontology. Moreover, if the conceptual model found for the ontology after the reverse engineering step shows some deficiencies, it may be improved through a restructuring step.

of our work was done at the knowledge level. This simplified the overall process of integration a lot.

We would also like to point out that in both cases there was no need to translate ontologies between different knowledge representation languages. 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; Russ *et al.*, 1999], translation is far from being a fully automatic process in the near future.

3.2 Integration activities

We are going to describe the most important activities that compose the ontology integration process. All integration activities assume that the ontology building activities are also performed, that is, the integration process does not substitute the ontology building process, it rather is a part of it.

Identify the possibility of integration The framework being used to build the ontology should allow some kind of knowledge reuse. For instance, the Ontolingua Server [Farquhar *et al.*, 1996] maintains an ontology library and allows integration operations, such as inclusion or restriction. More general systems, such as KACTUS do not allow such kind of operations, but allow pre-existent ontologies to be imported and edited. In other cases, integration (or any kind of reuse) may involve rebuilding an ontology in a framework different from the one where the ontology is available. In some cases, this may be cost-effective, in others it may be more cost-effective to build a new ontology from scratch that perfectly meets present needs and purposes than to try to rebuild and adapt a pre-existent ontology.

Identify the modules in which the ontology can be divided into The modules (building blocks) needed to build the future ontology are identified, that is, in which subontologies should the future ontology be divided (in integration, the modules are obviously related to ontologies). Upper-level modules and domain modules have to be identified. [5]

Identify the assumptions and ontological commitments that each module should comply to The assumptions and ontological commitments [Gruber, 1995] are described in the conceptual model and in the specification requirements document of the future ontology. This is one of the activities where documentation of an ontology can be crucial to allow better, faster and easier reuse. The assumptions and ontological commitments of the building blocks should be compatible among themselves and should be compatible with the assumptions and ontological commitments found for the resulting ontology.

Identify what knowledge should be represented in each module At this stage, one is only trying to have an idea of what the modules that are going to compose the future ontology should "look like" in order to recognize whether available ontologies are adequate to be reused. At this stage one only identifies a list of essential concepts. The conceptual model of the ontology and abstraction capabilities are used to produce such list.[6]

Identify candidate ontologies that could be used as modules This is subdivided into: (1) *finding available ontologies*, and (2) *choosing from the available ontologies which ones are possible candidates to be integrated*. To find possible ontologies one uses ontology sources. Since available ontologies are mainly implemented ones one should look for them in ontology libraries, as for instance, in the Ontolingua Server[7] for ontologies written in Ontolingua, in Ontosaurus[8] [Swartout *et al.*, 1997] for ontologies implemented in Loom [MacGregor, 1990a], or in the Cyc Server[9] for Cyc's upper-level ontology. Conceptualized or formalized ontologies are more difficult to find. Sometimes they are available in the literature or can be obtained by contacting ontology builders. However, not every ontology in a given subject will be appropriate to be reused. Some may lack some important concepts, etc. Therefore, from the available ontologies, one must chose those that satisfy a series of requirements. In the next section we discuss in detail how this choice is performed.

Get candidate ontologies in an adequate form This includes, not only, its knowledge level or implementation level *representations*, but also, all available *documentation*. As already discussed, one should prefer to work with the knowledge level representation of an ontology, if available. In some cases, this representation can be found in the literature (technical reports, books, thesis, etc.), or at least parts of it. Another possibility is contact ontology developers. However, in most cases, only the implementation level representation of an ontology is available, or is more easily available. Therefore, the *reengineering process* may be applied using the particular framework that was adopted to design the resulting ontology. If the ontology is not available (either at the implementation or knowledge level), one can still try to reconstruct it, or, at least, parts of it, using available documentation. While getting the implementation level representation of an ontology, if the ontology is not written in the adequate language (the language that was chosen to represent the resulting ontology) a knowledge *translation process* must take place. There are only a few translation attempts. In general, there are not many translators available, their technology is still immature and improving existing translators is a rather difficult task. In [Uschold *et al.*, 1998] the translation was done by hand and the conclusion was that this process is far from being a fully automatic process in the near future. Automatic translators are still at draft level [Russ *et al.*, 1999], therefore a lot of human intervention is needed to improve ontology translated versions. If translators are available they should be used to produce initial versions. Then, these initial versions should be improved by hand. Translators between different knowledge level representation languages are currently not available. The translation process is, in general, complex. It is important that, if the ontology includes other ontologies, one should also get the included ontolo-

[5] Representation ontologies are chosen in any ontology building process. Therefore, they are not specifically addressed here.

[6] At later stages one will need to know to what level of detail should that knowledge be represented, which relations should organize (structure) the ontology, and it would be helpful to know how it should be represented (concept, relation, etc.).

[7] `http://WWW-KSL-SVC.stanford.edu:5915`

[8] `http://www.isi.edu/isd/ontosaurus.html`

[9] `http://www.cyc.com`

gies. When reusing/using one ontology one must understand it fully, which includes every definition of every knowledge piece represented in the ontology (directly or indirectly). Included ontologies are a hidden part of the ontology. Knowledge pieces from the included ontologies can be used in the definitions of the ontology, therefore, in order to understand the ontology and know what is meant by one knowledge piece that comes from an included ontology one must have access to it and its definition or its technical documentation.

Study and analysis of candidate ontologies This includes two important activities: (1) *technical evaluation* of the candidate ontologies by *domain experts* through specialized criteria *oriented to integration* and (2) *user assessment* of the candidate ontologies by *ontologists* through specialized criteria *oriented to integration*. 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. In the next section we discuss the criteria to be used.

Choosing the most adequate source ontologies to be reused At this stage, and given the study and analysis of candidate ontologies performed by domain experts and ontologists, the final choices must be made. Among the chosen candidate ontologies that were technically evaluated and user assessed for integration one has to choose the ontology (or set of ontologies) that best suit our needs and purpose, or that can more easily or better be adapted to them. The ontology(ies) chosen to be reused may lack knowledge, may require that some knowledge is removed, etc., that is, it(they) 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. 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 reused ontologies are compatible among themselves, namely in what concerns the overall coherence. Sometimes, one can choose more than one ontology in a given subject if each one focuses different points of view of that subject. In the next section we go into the details of this choice.

Integrate knowledge All 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. When this part of the process ends, that is the appropriate ontologies to be reused in one particular integration process are found, we must integrate the knowledge of those ontologies. For that, one needs *integration operations* and *integration oriented design criteria*. 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. These can be viewed as composing, combining, modifying or assembling operations. 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). Design criteria guide the application of integration operations so that the resulting ontology has an adequate design and is of quality. In the next section we discuss the integration operations that were found useful in our integration experiences and the design criteria that guided their application.

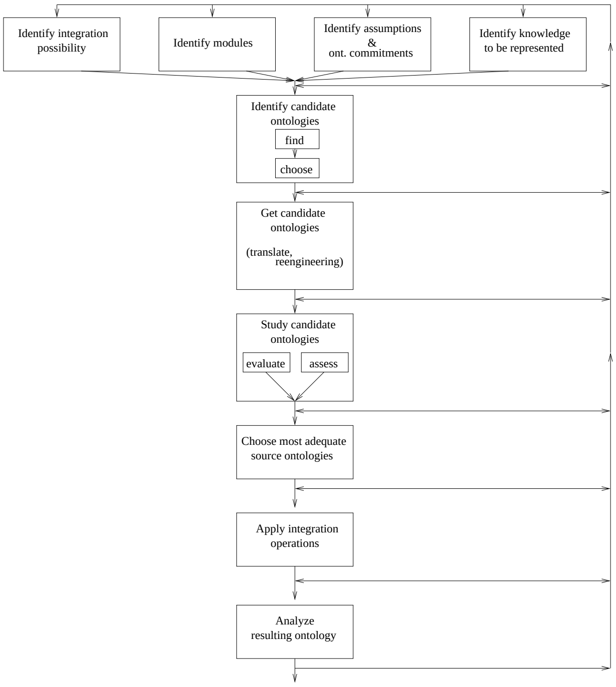
Identify integration possibility
Identify modules
Identify assumptions & ont. commitments
Identify knowledge to be represented
Identify candidate ontologies
find
choose
Get candidate ontologies
(translate, reengineering)
Study candidate ontologies
evaluate
assess
Choose most adequate source ontologies
Apply integration operations
Analyze resulting ontology

Figure 1: The integration process

Analyze resulting ontology After integration of knowledge one should evaluate and analyze the resulting ontology. Besides the usual criteria involved in evaluation of any ontology [Gómez-Pérez *et al.*, 1995] and the features that any ontology with an adequate design should comply to [Gruber, 1995] one should pay attention to specialized criteria that specifically analyzes whether the resulting ontology has enough quality. They are discussed in the next section.

3.3 Discussion

In Figure 1 we present the activities that compose the ontology integration process. Although ontology building and consequently ontology integration follows an evolving prototyping life cycle, some order must be followed. In general, the activities that compose the integration process tend to be performed following the order by which they were presented. However, some of the activities (and subactivities) to be performed before applying integration operations are interchangeable and some may be even performed in parallel. For instance, integration-oriented technical evaluation and user assessment of candidate ontologies. Moreover, the auxiliary subprocesses, reengineering and translation, may not occur in a particular integration process. 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

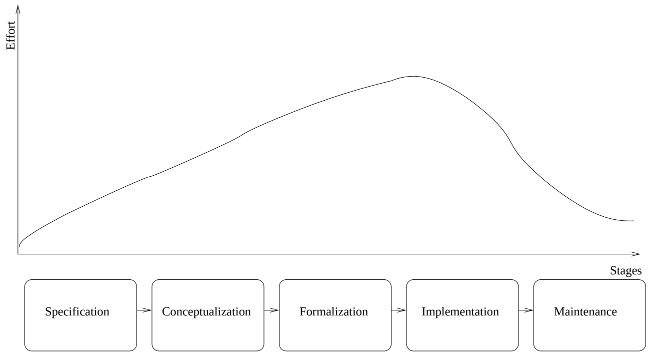

Figure 2: Integration effort along the ontology building process

the choice of the most adequate one remain essential activities to be performed. Finally, one can go back from any stage in the process to any other stage as entailed by the kind of life cycle. The important issue is that these activities are present in any integration process, although sometimes not explicitly or with different levels of importance and effort.

All activities, in particular those that precede application of integration operations, should be performed preferably in conceptualization or in formalization stages, that is, before implementation (some methodologies jump directly from conceptualization to implementation). However, if integration begins later in the ontology development life cycle, they still have to be performed. In both our integration experiences the framework that we used, ODE, automatically generated the implemented versions of the resulting ontologies. Therefore, we performed all integration activities during conceptualization and formalization stages. Using other frameworks may extend the process a bit. If the framework being used does not generate the implementation of the resulting ontology from the conceptual representations, after performing all activities at the knowledge level, the implemented versions of the chosen ontologies must be obtained and then one must apply the already determined sequence of integration operations in order to build the implemented version of the resulting ontology. In this case, only two activities (get ontologies and apply integration operations) had to be performed at the implementation level. This particular process falls into a typical evolving prototyping life cycle.

One important aspect of integration is the fact that this process is included in the overall ontology building process. The relation between the integration process and the overall ontology building process is shown in Figure 2. In the case that an ontology adequate to be reused is not found one must build it from scratch using one of the available ontology building methodologies.

The integration effort grows from specification and conceptualization to formalization where it reaches its maximum. It begins to decrease during implementation. It should be noted that in our particular case, due to the particular framework that was used the integration effort during implementation was null. The integration effort is not null during maintenance since integrated ontologies may themselves change due to maintenance activities making it necessary (or desirable) to reapply the integration process.

4 A Methodology

In this section we present the methods, procedures and guidelines that we developed to perform the activities that form this process. They form a methodology to perform integration.

4.1 Choosing candidate ontologies

To *choose candidate ontologies* one analyzes a series of features.[10] At this stage of the ontology integration process one is not going to be very particular, fussy, about the ontology, since one does not want to leave out any possible candidate. Therefore, only a very general analysis is made. Some of those features are *strict requirements*:

1. domain
2. is the ontology available?
3. formalism paradigms in which the ontology is available
4. main assumptions and ontological commitments
5. main concepts represented

If the ontology does not have adequate values for these properties they cannot be considered for integration. Therefore, these properties are used to eliminate ontologies. Other features are *desirable requirements* or desirable information:

1. where is the ontology available?
2. at what level is the ontology available?
3. what kind of documentation is available (technical reports, articles, etc.)?
4. where is that documentation available?

If some of the properties have certain values, the ontology is a better candidate: if the knowledge level representation of an ontology is available, then this ontology is a better candidate since the reengineering process would not have to be performed, if the internal and external documentation is available, then the most relevant information about the construction and choices made during the construction of the ontology is available, but if only articles are available about the ontology, then it is likely that some of the choices are not explained. If all of the values of these properties are unknown, then the ontology will not be a candidate, that is, if one cannot find where the ontology and the documentation is available, one cannot reuse it, therefore, the ontology is not a candidate. However, if there is enough documentation available, then it may be possible to reconstruct the ontology, and if the ontology is available, then it may be possible to understand it, provided that the domain is common enough and the ontology is simple and not very large (and possibly after some knowledge acquisition).

One can use a very simple metric to combine these different features. If strict requirements do not have adequate values, the ontology is eliminated. If desirable requirements

[10] Here we only describe the most important features involved in this choice. They are all organized into a taxonomy.

have appropriate values, then the ontology is a better candidate. If not, they are a worse candidate. If none of the desirable requirements have appropriate values, then the ontology is not a candidate. One does not want to eliminate any possible candidate at this stage of the integration process, only those that are of no use at all.

If, in a particular integration process, other features should be taken into consideration while choosing candidate ontologies, the metrics can be easily updated to take into account those new features. One only has to decide whether they are strict or desirable requirements The advantage of the flexibility of this metric is the fact that it can be better adapted to integration processes that should take into account particular features during the choice of one ontology. In particular, this kind of changes can narrow down the possible ontologies to choose from, if one introduces more strict requirements. For instance, one can impose the condition that only already evaluated ontologies should be considered as candidates. In that case, one should add this feature as a strict requirement. If one only wishes to prefer already evaluated ontologies, then this feature should be added as a desirable requirement.

4.2 Study and analysis of candidate ontologies

To *technically evaluate candidate ontologies* the domain experts should analyze the ontology paying special attention to [Pinto, 1999a; Pinto and Martins, 2000]:

- what knowledge is missing (concepts, classification criteria, relations, etc),
- what knowledge should be removed,
- which knowledge should be relocated,
- which knowledge sources changes should be performed,
- which documentation changes should be performed,
- which terminology changes should be performed,
- which definition changes should be made,
- which practices changes should be made.

Since domain experts usually find the languages used to implement ontologies difficult to understand [Fernández *et al.*, 1999], they should preferably be given a knowledge level representation of the ontology.

To *user assess candidate ontologies* the ontologists should analyze the ontology paying special attention to [Pinto, 1999a; Pinto and Martins, 2000]:

- the overall structure of the ontology (one hierarchy, several hierarchies, a graph, etc.) to assess whether the ontology has an adequate (and preferably well-balanced) structure, adequate and enough modules, adequate and enough specialization of concepts, adequate and enough diversity, similar concepts are represented closer whereas less similar concepts are represented further apart, knowledge is correctly "placed" in the structure so that inheritance mechanisms can infer appropriate knowledge from the ontology, etc;
- the distinctions (classification criteria made of the concepts described in the ontology) upon which the ontology is built to assess whether they are relevant and exactly the ones (quantity and quality) required;
- the relation used to structure knowledge[11] in the ontology to assess whether it is the required one;
- the naming convention rules used to assess whether they ease and promote reuse;
- the quality of the definitions (do they follow unified patterns, are simple, clear, concise, consistent, complete, correct —lexically and syntactically—, precise and accurate);
- the quality of the documentation of the ontology,
- the knowledge pieces represented (or included) are the ones that should be represented and all appropriate knowledge pieces are represented, etc.

Both domain experts and ontologists should evaluate and assess all and the whole of possible candidate ontologies. In [Pinto and Martins, 2000] a detailed discussion about the sets of integration oriented evaluation and assessment criteria can be found.

4.3 Choosing source ontologies

Choosing source ontologies is a rather complex multi-criteria choice where a lot of different aspects are involved. It is a much more complex choice than choosing candidate ontologies. For this reason, we propose that the task of choosing source ontologies should be divided into *two stages*.

First stage

In the *first stage* one tries to find which candidate ontologies are best suited to be integrated. Domain expert and ontologist analyses are crucial in this process. We propose that candidate ontologies should be analyzed according to a taxonomy of features, Figure 3.

General features give general information about the ontology. It is important that the ontology is of an adequate type, (*general or domain*). Depending on the *formality* [Uschold and Gruninger, 1996] of the resulting ontology one may integrate different kinds of ontologies. *Development status* gives information about the degree of readiness of an ontology to be reused (intended, on-going, toy example, implemented, mature). A toy example will only have representative knowledge pieces represented. An implemented ontology can be a good candidate provided that it has been carefully built or it has been evaluated. A mature ontology used in applications is a good candidate. This ontology should be a more or less stable ontology (provided that the domain does not evolve very rapidly).

Development features are related to how the ontology was built. The *quality of knowledge sources* and *adequacy of knowledge acquisition practices* are analyzed during the domain expert integration-driven technical evaluation. It is important that the ontology is *maintained*. One interesting finding about ontologies is the fact that they evolve, are "living", since their domains also evolve. Therefore, if they are maintained, it is most likely that they are updated. If they are maintained, it is important to know how maintenance is performed. Maintenance policies differ in *who* changes the

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

- general
  - generality
  - formality
  - development status
- development
  - knowledge acquisition
    - quality of knowledge sources
    - adequacy of knowledge acquisition practices
  - maintenance
    - is it maintained?
    - who does maintenance?
    - how is maintenance done?
  - documentation
    - quality of the documentation available
    - is the available documentation complete?
  - implementation
    - language issues
      - language(s) in which it is available
      - translators: are there translators? for which languages? quality of those translators
      - properties needed of the KR system in which it is built
- content
  - level of detail
  - modularity
  - adequacy from the domain expert point of view
  - adequacy from the ontologist point of view

Figure 3: Features for choosing source ontologies, first stage

ontology (can anybody change the ontology, or only authorized personnel?) and *how* those changes are performed (is the ontology changed regardless of people that built it, use it or reuse it? are the suggestions of change previously discussed among those groups? is there any attempt to reach a consensus between those groups? is there a special board that decides upon suggestions for changes?). It is important that the *documentation* has enough *quality* (it is clear, it adequately describes the domain, the ontology, the alternative representations of that ontology and which alternatives were preferred) and is *complete* (the ontology is completely described).

The language in which the ontology is represented is a rather important issue. If the ontology is available in the required language the task is greatly simplified. Although translation of ontologies is an important activity in integration, the overall effort of building the ontology can be considerably lessened if we avoid it. Therefore, it is important to know in which *languages* the ontology is available, whether *translators* from those languages are available, *for which languages?* those translators are available and their *quality*. It is also important to know which *reasoning capabilities* are needed by the ontology from the knowledge representation system where it is implemented, in order to know whether the ontology can be represented under a different knowledge representation system. Even if translators are available, one may not be sure of the possibility of full translation between different knowledge representation systems. For instance, while translating an ontology represented in first order logic into a pure frame system, if axioms are represented, they are lost. Therefore, one needs to know, among other issues:

- *formalism paradigm* (frames, semantic networks, description logics, etc.),
- which *inference mechanisms* are needed (general purpose, automated concept classifier [MacGregor, 1990b], inheritance,[12] monotonic vs modal vs nonmonotonic),
- whether *contexts* are required.

Content features give information about what is represented in the ontology and how that knowledge is represented. One needs to know whether the ontology has an adequate *level of detail*, that is, enough intermediate concepts are represented between two arbitrary concepts. One also needs to know *which concepts are represented in which modules*.

Under the feature *adequacy from the domain expert point of view* several analyses are made: does the content of the ontology include most of the relevant knowledge pieces of the domain? is the terminology adequate? are the definitions adopted correct and widely accepted? is the ontology complete in relation to present needs (at least, one needs to know what important knowledge pieces are missing)? is there superfluous knowledge that should be removed from the ontology while integrating it?

Under the feature *adequacy from the ontologist point of view* several analyses are made: are the basic distinctions represented in the ontology appropriate? does the ontology have an adequate structure? is the ontology structured according to appropriate relations? are needed knowledge pieces represented (this covers issues like ”are the appropriate relations represented?”, “are certain key concepts represented?”)? are those knowledge pieces adequately represented (this covers issues like fidelity, minimal encoding bias, correction, coherence, granularity, conciseness, efficiency in terms of time and space[13])? do they follow adequate naming convention rules? can missing knowledge pieces be added to the ontology without sacrificing coherence and clarity (this covers issues like extendible)? is the ontology clear?

The preponderant parts in this choice are played by the adequacy analyses that domain experts and ontologists have made of the candidate ontologies.

Since this choice is rather complex, simple metrics as the ones proposed to choose candidate ontologies are rather limited. The development of accurate metrics is an important open research area in the OE field.

After the first stage, one has chosen one possible set of ontologies to be integrated. It may be possible to have more than one ontology about one particular domain in that set. Those different ontologies represent knowledge about the same domain from different perspectives. Those different perspec-

[12] Which kind? defeasible, strict, mixed; credulous vs skeptical; on-path vs off-path; bottom-up vs top-down.

[13] It is important to know if we are not reusing an ontology that is not going to meet our needs and the means that we currently have at our disposal.

- content
  - completeness
  - compatibility
    - terminology of common concepts
    - definitions of common concepts

Figure 4: Features for choosing source ontologies, second stage

tives should have been found important to be present in the resulting ontology (there should not be duplicated knowledge represented in the resulting ontology). However, the chosen ontologies may not be compatible among themselves.

Second stage

In the *second stage* one tackles compatibility and completeness of possibly chosen ontologies in relation to the desired resulting ontology, Figure 4.

If the ontologies which are possibly going to be chosen to be integrated are not coherent in what concerns the terminology used and the definitions of the concepts that are common to more than one ontology, then they are not *compatible* and, therefore, cannot be assembled. Sometimes the same concept is named differently in different ontologies. In the resulting ontology one concept only has one denomination, therefore one must be adopted. If one concept has the same definition in all chosen ontologies but different denominations, then a change in terminology can solve the problem. All definitions involving the renamed concept have to be checked and revised accordingly. Sometimes different ontologies adopt different definitions for the same concept. One cannot have this kind of inconsistencies in the resulting ontology. One definition should be chosen and adopted all over. It is more difficult to ensure that the same definition can be adopted by all integrated ontologies. A thorough analysis of all ontologies where one particular concept has a different definition from the adopted one has to be made. It is obvious that only a coherent set of ontologies should be considered for integration purposes.

If chosen ontologies are not *complete*, that is, they do not comprehend all the ontology that has to be built, then this piece of information must be known so that missing knowledge pieces are built from scratch and added or another compatible ontology that contains those knowledge pieces is integrated.

So, although the problem of lack of completeness has to be known, it is not as problematic as lack of coherence. Since one of the issues involved in the domain expert analysis is missing knowledge, one can check whether it is not represented in another ontology about the same domain that is also (or can also be) integrated. However, if chosen ontologies are not compatible among themselves, then this may imply choosing another possible set of ontologies by combining candidate ontologies into a different set, or it may imply building ontologies from scratch (if none of the candidate ontologies adopts the adequate terminology and definitions, or profound changes have to be made to them in order to integrate them).

The problem of choosing the appropriate set of source ontologies is also rather complex. From the set of candidate ontologies, a coherent and adequate subset must be found that is as close as possible to the resulting ontology. Once again, the ontologies in that set may not be perfect candidates. As long as the changes to be made are not very extensive it is more cost effective to reuse the ontology. This analysis has to be performed on a case by case basis. If it is more cost effective to build the ontology from scratch, then existing ontology building methodologies can be used to build an ontology that perfectly suits our needs. If not, ontologies should be reused and integration operations applied so that adequate changes transform the ontologies into perfect candidates.

The result of this activity is a set of ontologies that can and should be assembled together, a description of lacking knowledge that is going to be built from scratch and included in the resulting ontology (since none of the chosen ontologies has it and that knowledge has been identified as essential knowledge that must exist in the resulting ontology) and a description of the changes that should be performed to the integrated ontologies so that they can be perfect candidates and successfully reused (which is the starting point for the application of the integration operations).

4.4 Integration of knowledge

To *integrate knowledge* one needs integration operations and design criteria to guide their application. Sometimes the adaptation of source ontologies may require restructuring activities similar to those that are performed in reengineering processes. Moreover, it may require introduction/removal of knowledge pieces, correction and improvement of the definitions, terminology and documentation of the knowledge pieces represented in the ontology, etc. These adaptations transform the chosen ontology (whole of it) into the needed ontology. In [Farquhar *et al.*, 1997; Borst, 1997; Pinto and Martins, 2000; Pinto, 1999a] initial sets of integration operations are proposed. *Integration operations* can be divided into two groups: basic and non-basic. While the former operations can be algebraically specified the latter can be defined from the former but are custom-tailored operations to be defined in a case by case basis. We have developed an algebraic specification of 39 basic integration operations and specified how 12 non-basic operations can be defined from the previous ones. They are described in [Pinto, 1999b]. We identified a set of *criteria* to guide integration of knowledge: modularize, specialize, diversify each hierarchy, minimize the semantic distance between sibling concepts, maximize relationships between taxonomies and standardize names of relations. They are described in detail in [Arpirez-Vega *et al.*, 1998].

4.5 Analysis of resulting ontology

To *analyze the resulting ontology* one uses a set of features. Besides having an adequate design according to the set of features proposed in [Gruber, 1995][14] and compliance with evaluation criteria [Gómez-Pérez *et al.*, 1995;

[14] Clarity, coherence, extendibility, minimal encoding bias and minimal ontological commitment.

Gómez-Pérez, 1996; 1999][15], one should pay attention to whether the ontology has a *regular level of detail all over*. By regular level of detail we mean that there are no ”islands” of exaggerated level of detail and other parts with an adequate one. 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. It should also be noted that the other features involved in evaluation and design criteria are analyzed in relation to the resulting ontology, for instance, the resulting ontology should be consistent and coherent all over (although composed by knowledge from different ontologies).

5 Conclusions

In this article we presented the characterization of the ontology integration process. The activities that compose this process are described. The most important activities that form this process include: finding and choosing candidate ontologies, integration oriented evaluation and assessment of candidate ontologies, choosing adequate source ontologies to be integrated, application of integration operations to integrate knowledge and analysis of the resulting ontology. We describe the methods developed to perform these activities. They provide support and guidance to the activities that compose the integration process. They form an integration methodology.

The advantages of the proposed integration methodology are a direct consequence of its generality. One of the advantages of our integration methodology is the fact that it *can be used with different methodologies to build ontologies from scratch*. The only assumption made by this methodology is that knowledge should be represented at the knowledge level.

Special emphasis is given to the *quality* of the ontologies involved in a particular integration process. There are two cases in what regards the ontologies that are reused: (1) they are available at ontology libraries and were built by others or (2) they were built by us. Our methodology proposes that all reused ontologies should be evaluated by domain experts from a technical point of view and assessed by ontologists (more precisely by the ontologists that are going to play the role of integrators) from a user point of view. Integration-oriented technical evaluation and user assessment criteria assure that reused ontologies have enough technical quality to be used in the process. The analysis of the resulting ontology assures that the resulting ontology has enough quality to be made available and (re)used.

References

[Amaya, 1998] M. Dolores Rojas Amaya. Ontologia de Iones Monoatómicos en Variables Fisicas del Medio Ambiente. Proyecto Fin de Carrera, Fac. de Informática, UPM, 1998.

[Arpirez-Vega *et al.*, 1998] J. Arpirez-Vega, A. Gomez-Perez, A. Lozano-Tello, and H. Sofia Pinto. (ONTO)^{2}Agent: An Ontology-Based WWW Broker to Select Ontologies. In *Proceedings of ECAI98's Workshop on Application of Ontologies and Problem Solving Methods*, pages 16–24, 1998.

[Arpirez-Vega *et al.*, 2000] J. Arpirez-Vega, A. Gomez-Perez, A. Lozano-Tello, and H. Sofia Pinto. Reference Ontology and (ONTO)^{2}Agent: the Ontology Yellow Pages. *Knowledge and Information Systems*, 2(4):387–412, 2000.

[Benjamins and Fensel, 1998] Richard Benjamins and Dieter Fensel. The Ontological Engineering Initiative (KA)^2. In Nicola Guarino, editor, *Formal Ontology in Information Systems*, pages 287–301. IOS Press, 1998.

[Benjamins *et al.*, 1999] Richard Benjamins, Dieter Fensel, Stefan Decker, and Asunción Gómez-Pérez. (KA)^2: Building Ontologies for the Internet, a Mid Term Report. *International Journal of Human Computer Studies*, 51:687–712, 1999.

[Blázquez *et al.*, 1998] M. Blázquez, Mariano Fernández, J. M. García-Pinar, and Asunción Goméz-Pérez. Building Ontologies at the Knowledge Level Using the Ontology Design Environment. In *Proceedings of the Knowledge Acquisition Workshop, KAW98*, 1998.

[Borst, 1997] Pim Borst. *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. PhD thesis, Tweente University, 1997.

[Chandrasekaran *et al.*, 1999] B. Chandrasekaran, J.R. Josephson, and V. Richard Benjamins. Ontologies: What are they? Why do we need them? *IEEE Expert (Intelligent Systems and Their Applications)*, 14(1):20–26, 1999.

[Farquhar *et al.*, 1996] Adam Farquhar, Richard Fikes, and James Rice. The Ontolingua Server: A Tool for Collaborative Ontology Construction. In *Proceedings of the Knowledge Acquisition Workshop, KAW96*, 1996.

[Farquhar *et al.*, 1997] Adam Farquhar, Richard Fikes, and James Rice. Tools for Assembling Modular Ontologies in Ontolingua. In *AAAI97 Proceedings*, pages 436–441. AAAI Press, 1997.

[Fernández *et al.*, 1997] Mariano Fernández, Asunción Gómez-Pérez, and N. Juristo. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. In *Proceedings of AAAI97 Spring Symposium Series, Workshop on Ontological Engineering*, pages 33–40, 1997.

[Fernández *et al.*, 1999] Mariano Fernández, Asunción Gómez-Pérez, Alexandro Pazos Sierra, and Juan Pazos Sierra. Building a Chemical Ontology Using METHONTOLOGY and the Ontology Design Environment. *IEEE Expert (Intelligent Systems and Their Applications)*, 14(1):37–46, 1999.

[Gangemi *et al.*, 1998] Aldo Gangemi, Domenico M. Pisanelli, and Geri Steve. Ontology Integration: Experiences with Medical Terminologies. In Nicola Guarino, editor, *Formal Ontology in Information Systems*, pages 163–178. IOS Press, 1998.

[15] Correctness –lexically and syntactically–, completeness, conciseness, consistency, expandability, sensitiveness and robustness.

[Gómez-Pérez and Rojas-Amaya, 1999] Asunción Gómez-Pérez and Dolores Rojas-Amaya. Ontological Reengineering for Reuse. In D. Fensel and R. Studer, editors, *Proceedings of the European Knowledge Acquisition Workshop, EKAW99*. Springer Verlag, 1999.

[Gómez-Pérez *et al.*, 1995] A. Gómez-Pérez, N. Juristo, and J. Pazos. Evaluation and Assessment of the Knowledge Sharing Technology. In N.J.I. Mars, editor, *Towards Very Large Knowledge Bases*, pages 289–296. IOS Press, 1995.

[Gómez-Pérez *et al.*, 1996] Asunción Gómez-Pérez, Mariano Fernández, and António J. de Vicente. Towards a Method to Conceptualize Domain Ontologies. In *Proceedings of ECAI96's Workshop on Ontological Engineering*, pages 41–52, 1996.

[Gómez-Pérez, 1996] Asunción Gómez-Pérez. Towards a Framework to Verify Knowledge Sharing Technology. *Expert Systems with Applications*, 11(4):519–529, 1996.

[Gómez-Pérez, 1999] Asunción Gómez-Pérez. Evaluation of Taxonomic Knowledge in Ontologies and Knowledge Bases. In *Proceedings of the Knowledge Acquisition Workshop, KAW99*, 1999.

[Gruber and Olsen, 1994] Thomas Gruber and G. R. Olsen. An Ontology for Engineering Mathematics. In J. Doyle, E. Sandewall, and P. Torasso, editors, *KR94 Proceedings*, pages 258–269. Morgan Kaufmann, 1994.

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

[Gruninger, 1996] Michael Gruninger. Designing and Evaluating Generic Ontologies. In *Proceedings of ECAI96's Workshop on Ontological Engineering*, pages 53–64, 1996.

[MacGregor, 1990a] Robert MacGregor. LOOM User Manual. Technical Report ISI/WP-22, USC/Information Sciences Institute, 1990.

[MacGregor, 1990b] Robert MacGregor. The Evolving Technology of Classification-Based Representation Systems. In John Sowa, editor, *Principles of Semantic Networks: Explorations in the Representation of Knowledge*, pages 385–400. Morgan Kaufman, 1990.

[McGuiness *et al.*, 2000] Deborah L. McGuiness, Richard Fikes, James Rice, and Steve Wilder. An Environment for Merging and Testing Large Ontologies. In Anthony Cohn, Fausto Giunchiglia, and Bart Selman, editors, *KR2000 Proceedings*, pages 483–493. Morgan Kaufmann, 2000.

[Newell, 1982] A. Newell. The Knowledge Level. *Artificial Intelligence*, 18(1):87–127, 1982.

[Noy and Musen, 1999] Natalya Fridman Noy and Mark A. Musen. An Algorithm for Merging and Aligning Ontologies: Automation and Tool Support. In *Proceedings of AAAI99's Workshop on Ontology Management, WS-99-13*, pages 17–27. AAAI Press, 1999.

[Noy and Musen, 2000] Natalya Fridman Noy and Mark A. Musen. PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment. In *AAAI2000 Proceedings*, pages 450–455. AAAI Press, 2000.

[Pinto and Martins, 2000] H. Sofia Pinto and J.P. Martins. Reusing Ontologies. In *Proceedings of AAAI 2000 Spring Symposium Series, Workshop on Bringing Knowledge to Business Processes, SS-00-03*, pages 77–84. AAAI Press, 2000.

[Pinto *et al.*, 1999] H. Sofia Pinto, A. Gómez-Pérez, and J. P. Martins. Some Issues on Ontology Integration. In *Proceedings of IJCAI99's Workshop on Ontologies and Problem Solving Methods: Lessons Learned and Future Trends*, pages 7.1–7.12, 1999.

[Pinto, 1999a] H. Sofia Pinto. Towards Ontology Reuse. In *Proceedings of AAAI99's Workshop on Ontology Management, WS-99-13*, pages 67–73. AAAI Press, 1999.

[Pinto, 1999b] H. Sofia Pinto. Towards operations to ontology integration. Technical Report GIA 99/02, Grupo de Inteligência Artificial do Instituto Superior Técnico, April 1999.

[Russ *et al.*, 1999] Thomas Russ, Andre Valente, Robert MacGregor, and William Swartout. Practical Experiences in Trading Off Ontology Usability and Reusability. In *Proceedings of the Knowledge Acquisition Workshop, KAW99*, 1999.

[Sowa, 2000] John Sowa. *Knowledge Representation: logical, philosophical and computational foundations*. Brooks/Cole, 2000.

[Swartout *et al.*, 1997] Bill Swartout, Ramesh Patil, Kevin Knight, and Tom Russ. Toward Distributed Use of Large-Scale Ontologies. In *Proceedings of AAAI97 Spring Symposium Series, Workshop on Ontological Engineering*, pages 138–148, 1997.

[Uschold and Gruninger, 1996] Mike Uschold and Michael Gruninger. Ontologies: Principles, Methods and Applications. *Knowledge Engineering Review*, 11(2), June 1996.

[Uschold and King, 1995] Mike Uschold and Martin King. Towards a Methodology for Building Ontologies. In *Proceedings of IJCAI95's Workshop on Basic Ontological Issues in Knowledge Sharing*, 1995.

[Uschold *et al.*, 1998] Mike Uschold, Mike Healy, Keith Williamson, Peter Clark, and Steven Woods. Ontology Reuse and Application. In Nicola Guarino, editor, *Formal Ontology in Information Systems*, pages 179–192. IOS Press, 1998.

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