COST Technical Committee "Transport and Urban Development"

COST Action C21
"Urban Ontologies for an improved communication in urban civil engineering projects" - TOWNTOLOGY Project
http://www.towntology.net/

Technical report n°1
Guidelines to build ontologies: A bibliographic study

Catherine Roussey (LIRIS, INSA de Lyon)

Version: 1
Preparation date: November 2005

Guidelines to build ontology: A Bibliographic study

Catherine Roussey

This report is based on [Gandon, 2002] study.

- Chapter 1 is a presentation of the object ontology, we propose some definition, follow up by some classification of ontology type. Thus, these classification enable us to propose a list of ontology component. Then a list a famous ontology is proposed.
- Chapter 2 propose some bibliography about ontology development.
- Chapter 3 is a description of the different stage of ontology development.

1 Ontology

1.1 Ontology definition

"The word ontology comes from the Greek ontos for being and logos for word. It is a relatively new term in the long history of philosophy, introduced by the 19th century German philosophers to distinguish the study of being as such from the study of various kinds of beings in the natural sciences. The more traditional term is Aristotle's word category (kathgoria), which he used for classifying anything that can be said or predicated about anything." [Sowa, 2000]

"An ontology is an explicit specification of a conceptualization" [Gruber 1993]

"An ontology is a formal and explicit specification of a share conceptualization" [Studer 1998]

"Conceptualization contains the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them [Genesereth & Nilsson, 1987] . A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly" [Gruber web site].

Explicite means that concept types and their usage constraints are explicitly defined.

Formal means machine understandable.

Share means concensual knowledge accepted by a group.

1.2 Ontology component

Different types of ontology exist and several classifications have been proposed. These classification are based on the characteristic of the ontology component.

1.2.1 Uschold and Gruninger Classification

[Uschold and Gruninger, 1996] proposed four types of ontology dependant of their formalization:

- **highly informal**: the ontology is expressed loosely in natural language;
- **semi-informal**: the ontology is expressed in a restricted and structured form of natural language, greatly increasing clarity by reducing ambiguity;
- **semi-formal**: the ontology is expressed in an artificial formally defined language;
- **rigorously formal**: the ontology is meticulously defined with formal semantics, theorems and proofs.

1.2.2 Ontoware classification

[http://cos.ontoware.org/] propose 3 different classifications of ontology:

Classification according to Purpose

o Application Ontology

Used in a specific application implementing a reasoner based on an ontology. The typical trade-off between expressiveness and decidability requires a limited representation formalism.

o Reference Ontology

Used during development time of applications for mutual understanding and explanation between (human or artificial) agents belonging to different communities, for establishing consensus in a community that needs to adopt a new term or simply for explaining the meaning of a term to somebody new to the community.

Classification according to Expressiveness

o Heavyweight Ontology

Heavyweight ontologies are extensively axiomatized and thus represent ontological commitment explicitly. The purpose of the axiomatization is to exclude terminological and conceptual ambiguities, due to unintended interpretations. Every heavyweight ontology can have a lightweight version. Many domain ontologies are heavyweight because they should support heavy reasoning (e.g., for integrating database schemata, or to drive complex corporate applications). As with all dimensions, the borderline between light and heavy weight is not clearly delimited.

o Lightweight Ontology

Lightweight ontologies are simple taxonomic structures of primitive or composite terms together with associated definitions. They are hardly axiomatized as the intended meaning of the terms used by the community is more or less known in advance by all members, and the ontology can be limited to those structural relationships among terms that are considered as relevant.

Classification according to Specificity

o Generic Ontology

The concepts defined by this layer are considered to be generic across many fields. Typically, generic ontologies (synonyms are "upper level" or "top-level" ontology) define concepts such as state, event, process, action, component etc.

o Core Ontology

Core ontologies define concepts which are generic across a set of domains. Therefore, they are situated in between the two extremes of generic and domain ontologies. The borderline between generic and core ontologies is not clearly defined because there is no exhaustive enumeration of fields and their conceptualizations. However, the distinction is intuitively meaningful and useful for building libraries.

o Domain Ontology

Domain ontologies express conceptualizations that are specific for a specific universe of discourse. The concepts in domain ontologies are often defined as specializations of concepts in the generic and core ontologies. The borderline between core and domain ontologies is not clearly defined because core ontologies intend to be generic within a domain. Thus, it is usually hard to make a clear cut between generic and core as well as between core and domain ontologies. A concept such as software component would be placed in a core ontology for application servers for reuse in every possible domain ontology we can think of. However, a concept such as enterprise bean might only be relevant in a specific J2EE setting.

1.2.3 Propose Classification

We can propose a distinction between lightweight / heayvyweight ontology according to their components.

Ontology components	Definition propose by the free dictionary *(http://www.thefreedictionary.com/*)		
Term or phrase	*a word or expression used for some particular thing. A term is the linguistic representation of a concept.*	Lightweight ontology	Heavyweight ontology
Natural language Definition	Informal definitions are human understandable.		
Concept, class or concept type	*an abstract or general idea inferred or derived from specific instances. For example concept contains the characteristics of a group of individuals objects.* *Something formed in the mind; a thought or notion.*		
Relation, relation type	*an abstraction belonging to or characteristic of two entities or parts together*		
Instance, individual	*an occurrence of something;* *an item of information that is representative of a type;*		
Relation instance	An occurrence of a relation type.		
Function	Function are used to compute some value associate to a concept (age for example)		
Axiom, fact	*a statement that affirms or denies something and is either true or false*		
Attribute, property	*an abstraction belonging to or characteristic of a concept*		
Formal definition	The definition of a concept in a formal language like description logics, conceptual graph, etc… this definition is machine understandable… the definition is an axiom.		
Rules	Rule are elementary component to reasoning and infert new knowledge. For example a rule is like IF X then Y.		

1.3 Ontology example

AAT: Art & Architecture Thesaurus to describe art, architecture, decorative arts, material culture, and archival materials.

Airport Codes: simple ontology containing Airport codes around the world.

ASBRU: provides an ontology for guideline-support tasks and the problem-solving methods in order to represent and to annotate clinical guidelines in standardised form.

Bibliographic Ontology: a sample bibliographic ontology, with data types taken from ISO standard.

FIPA Agent Communication Language: contains an ontology describing speech acts for artificial agents communication.

CHEMICALS: Ontology containing knowledge within the domain of chemical elements and crystalline structures.

CoreLex: an ontology for lexical semantic database and tagset for nouns, organised around systematic polysemy and underspecification.

The Dublin Core Element Set Ontology: ontology interoperable metadata standards and developing specialised metadata vocabularies for describing resources that enable more intelligent information discovery systems.

EngMath: mathematics engineering ontologies including ontologies for scalar quantities, vector quantities, and unary scalar functions.

Enterprise Ontology: The Enterprise Ontology is a collection of terms and definitions relevant to business enterprises. The ontology was developed in the Enterprise Project by the Artificial Intelligence Applications Institute at the University of Edinburgh with its partners: IBM, Lloyd's Register, Logica UK Limited, and Unilever. The project was support by the UK's Department of Trade and Industry under the Intelligent Systems Integration Programme (project IED4/1/8032). Conceptually, the Enterprise Ontology it is divided into a number of main sections: activities and processes, organisation, strategy and marketing.

Gene Ontology: A dynamic controlled vocabulary that can be applied to all eukaryotes even as knowledge of gene and protein roles in cells is accumulating and changing. The three organising principles of this ontology are molecular function, biological process and cellular component.

Gentology: genealogy ontology for data interchange between different applications.

$(KA)^2$ The Knowledge Annotation Initiative of the Knowledge Acquisition Community develop an ontology that models the knowledge acquisition community (its researchers, topics, products, etc.). http://ksi.cpsc.ucalgary.ca/KAW/KAW98/benjamins2/

Knowledge Representation Ontology: from the book Knowledge Representation by John F. Sowa, this ontology proposes a top level for knowledge representation based on basic categories and distinctions that have been derived from a variety of sources in logic, linguistics, philosophy, and artificial intelligence.

Open Cyc: an upper ontology for all of human consensus reality i.e., 6000 concepts of common knowledge. http://www.cyc.com

ONTOGEO: This research group study ontology development, use and implementation in GIS system. They propose a Greek cadastre ontology made in Protégé-2000, (http://ontogeo.ntua.gr/)

PLANET: Planet is a reusable ontology for representing plans that is designed to accommodate a diverse range of real-world plans, both manually and automatically created.

ProPer: ontology to manage skills and competencies of people.

SENSUS: concept taxonomies for natural language processing. Sensus is used in the Pangloss Machine Translation systems, the Penman sentence generation system, and eventually other systems as appropriate. SENSUS is an extension and reorganization of WordNet. at the top level, we have added nodes from the Penman Upper Model, and the major branches of WordNet have been rearranged to fit. http://www.isi.edu/natural-language/projects/ONTOLOGIES.html

SurveyOntology: ontology used to describe large questionnaires; it was developed for clients at the census bureau and department of labour.

TOVE: The goal of the TOronto Virtual Enterprise project is to create a data model that provides a shared terminology for the enterprise, defines the meaning of each term in a precise and as unambiguous manner as possible, implements the semantics in a set of axioms, and defines a symbology for depicting a the concepts.

UMDL Ontology: ontology for describing digital library content.

UMLS: the Unified Medical Language System provides a biomedical vocabulary from disparate sources such as clinical terminologies, drug sources, vocabularies in different languages, and clinical terminologies.

Wordnet, Eurowordnet : WordNet® is an online lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. Different relations link the synonym sets. See http://www.globawordnet.org

http://swoogle.umbc.edu/ is an research engine available for ontology written in OWL formal language…

2 Several method of ontology design

This chapter list some papers describing some ontology development method.

2.1 State-of-the-art of ontology methodology.

Several papers or books propose some state-of-the-art about ontology development methodology.

- SIGSEMIS Bulletin propose a column about methodologies for the semantic web including a state-of-the-art of ontology methodology see for example the SIGSEMIS Bulletin Vol 1, Issue 2 (July 2004) available http://www.sigsemis.org
- Steffen Staab, Rudi Studer (Eds.): Handbook on Ontologies. International Handbooks on Information Systems Springer 2004, ISBN 3-540-40834-7
- Oscar Corcho Mariano Fernández-López Asunción Gómez-Pérez Methodologies, tools and languages for building ontologies: where is their meeting point?. Data & Knowledge Engineering archive, Volume 46 , Issue 1 (July 2003)
- M. Fernández-López, Overview of Methodologies for Building Ontologies, in: IJCAI99 Workshop on Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends, Stockholm, 1999.

2.2 Methodology by Uschold and King

- Ontologies: Principles, Methods and Applications - Uschold, Grüninger (1996) Knowledge Engineering Review 11(2), 1996.
- Building Ontologies: Towards a Unified Methodology - Uschold (1996) 16th Annual Conf. of the British Computer Society Specialist Group on Expert Systems
- Towards a Methodology for Building Ontologies (1995) Mike Uschold, Martin King

2.3 TOronto Virtual Enterprise (TOVE) Methodology

URL: http://www.eil.utoronto.ca/enterprise-modelling/entmethod/index.html

People: Gruninger M, Fox MS

„*As part of the TOVE project, we set out to design a methodology for creating ontologies. The methodology is composed of the following steps:*

- *Define a set of Motivating Scenarios*
- *Define a set of Informal Competency Questions that the ontology must answers in order to support the motivating scenarios*
- *Using First-Order Logic, define the Terminology of the ontology*
- *Formally redefine the Competency Questions using the terminoloyg and first-order logic*
- *Define the semantics and constraints on the terminology using first-order logic*

The unique contribution of this work is the introduction of competency questions as a basis of defining the scope of an ontology. " [TOVE website]

- Methodology for the Design and Evaluation of Ontologies (1995) Michael Grüninger, Mark S. Fox IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing, April 13, 1995

2.4 Sensus methodology

url: http://www.isi.edu/~hovy/

People: Eduard H Hovy, Dr, Patrick Pantel, Mr. Michael Fleischman, Mr. Andrew Philpot, and Dr. Jerry Hobbs

"*The development of large concept taxonomies/ontologies through a combination of merging* *together existing ontologies, adding to the knowledge by extracting information from online text and enriching the interdependency relations by extracting information from dictionaries. Omega, our current ontology (2003–), contains over 120,000 concept terms and several million instances, in addition to various other information, acquired from a variety of sources, including Princeton's WordNet, NMSU's Mikrokosmos, and ISI's previous ontology SENSUS (1996–2000).* " [web site of EH Hovy]

- Eduard H. Hovy: Methodologies for the Reliable Construction of Ontological Knowledge. ICCS 2005: 91-106
- Book chapter: Hovy, E.H. 2002. Comparing Sets of Semantic Relations in Ontologies. In R. Green, C.A. Bean, and S.H. Myaeng (eds), The Semantics of Relationships: An Interdisciplinary Perspective, pp. 91–110. Dordrecht: Kluwer.

2.5 Methontology

Tools: ODE (Ontology Design Environment), webODE

people: Mariano Fernandez, Asuncion Gomez-Perez

"*the most famous ontology design environment… try to define the needed activity that people need to carry out when building an ontology*" [Sigsemis bulletin Vol1 No 2 july 2004]

- Mariano Fernández López, Asunción Gómez-Pérez, Juan Pazos Sierra, Alejandro Pazos Sierra. 1999. Building a Chemical Ontology Using Methontology and the Ontology Design Environment. IEEE Intelligent Systems, January/February 1999 (Vol. 14, No. 1).
- Fernández, M. Overview of Methodologies for Building Ontologies. Workshop on Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends. (IJCAI99). August. 1999.
- M. Blázquez, M. Fernández, J. M. García-Pinar, A. Gómez-Pérez. 1998. Building Ontologies at the Knowledge Level Using the Ontology Design Environment. Knowledge Acquisition Conference, 1998 (KAW-98).
- Gómez-Pérez, A.; Fernández, M.; de Vicente, A. (1996). Towards a Method to Conceptualize Domain Ontologies. Workshop on Ontological Engineering. ECAI'96. Budapest, Hungary. PP: 41-52.
- Gómez-Pérez. A. (1998). Knowledge Sharing and Reuse. The Handbook of Applied Expert Systems. Edited by Liebowitz. CRC.
- Mariano Fernandez, Asuncion Gomez-Perez, 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
- Fernandéz, Mariano; Gómez-Pérez, Asunción; and Juristo, Natalia. 1997. METHONTOLOGY: From Ontological Art to Ontological Engineering. Workshop on Ontological Engineering. Spring Symposium Series. AAAI97 Stanford.

2.6 Kactus methodology

URL: http://hcs.science.uva.nl/projects/NewKACTUS/home.html

Tools: VOID the KACTUS toolkit

People: B. J. Wielinga, A. Th. Schreiber, J. M. Akkermans

"KACTUS stands for modelling Knowledge About Complex Technical systems for multiple USe. It is an European ESPRIT-iii project aiming at the development of a methodology for the reuse of knowledge about technical systems during their life-cycle. This implies using the same knowledge base for design, diagnosis, operation, maintenance, redesign, instruction, et cetera. Reuse will be achieved by giving these knowledge bases an explicit structure (often called an ontology)."

- Bernaras, I. Laresgoiti, J. Corera, Building and reusing ontologies for electrical network applications, in: Proc. European Conference on Artificial Intelligence (ECAI'96), Budapest, Hungary, 1996, pp. 298-302.
- B. J. Wielinga and A. Th. Schreiber. Conceptual modelling of large reusable knowledge bases. In K. von Luck and H. Marburger, editors, Management and Processing of Complex Data Structures, volume 777 of Lecture Notes in Computer Science, pages 181-200, Berlin, Germany, 1994. Springer Verlag.
- A. Th. Schreiber, B. J. Wielinga, J. M. Akkermans, W. Van de Velde, and A. Anjewierden. CML: The CommonKADS conceptual modelling language. In L. Steels, A. Th. Schreiber, and W. Van de Velde, editors, A Future for Knowledge Acquisition. Proceedings of the 8th European Knowledge Acquisition Workshop EKAW'94, volume 867 of Lecture Notes in Artificial Intelligence, pages 1-25, Berlin/Heidelberg, September 1994. Springer-Verlag.
- B. J. Wielinga, J. M. Akkermans, and A. Th. Schreiber. A formal analysis of parametric design problem solving. In B. R. Gaines and M. A. Musen, editors, Proceedings of the 8th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop, volume II, pages 37-1 -- 37-15, Alberta, Canada, 1995. SRDG Publications, University of Calgary.
- A. Th. Schreiber, B. J. Wielinga, and W. H. J Jansweijer. The KACTUS view on the 'O' word. Technical Report, ESPRIT Project 8145 KACTUS, University of Amsterdam, 1995. Submitted for publication.
- A. Th. Schreiber and P. Terpstra. Sisyohus-VT: A CommonKADS solution. Technical Report, ESPRIT Project 8145 KACTUS, University of Amsterdam, 1995. Submitted for publication.

2.7 DOLCE : a Descriptive Ontology for Linguistic and Cognitive Engineering

The main author's idea is to develop a library of ontologies described using first-order logic : WonderWeb - Library of Foundation Ontologies. The specific assumptions adopted for the DOLCE module are first introduced informally, along with the basic categories, functions, and relations.

url: www.loa-cnr.it/DOLCE.html

tools: ONION, Ontoclean

People: Nicola Guarino, Borgo S,

- Gangemi, A., Catenacci, C., Battaglia, M. Inflammation Ontology Design Pattern: an Exercise in Building a Core Biomedical Ontology with Descriptions and Situations. In D.M. Pisanelli (ed.) Ontologies in Medicine, IOS Press, Amsterdam, Berlin, 2004, pp. 64-80
- Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A WonderWeb DeliverableD18; Ontology Library (final) (ver. 1.0, 31-12-2003)
- Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., Schneider, L. WonderWeb Deliverable D17. The WonderWeb Library of Foundational Ontologies and the DOLCE ontology. Preliminary Report (ver. 2.0, 15-08-2002)

3 Description of the ontology development

"*The design of an ontology is an iterative maturation process. This means the ontology will come to full development, becoming mature, by evolving through intermediate states to reach a desired state.*

As soon as the ontology becomes large, the ontology engineering process has to be considered as a project, and therefore, project management methods must be applied. [Fernandez et al., 1997] recognised that planning and specification are important activities. The authors give the activities to be done during the ontology development process (Figure 8): planning, specifying, acquiring knowledge, conceptualising, formalising, integrating, implementing, evaluating, documenting, and maintaining.

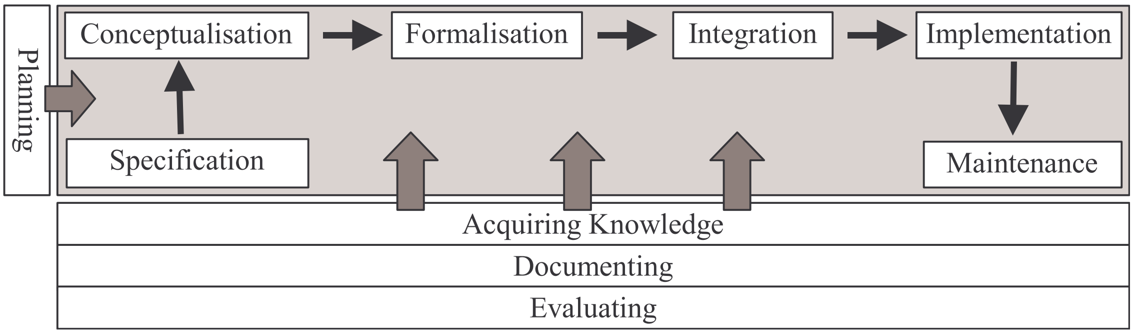

Figure 8 States and activities in the ontology lifecycle [Fernandez et al., 1997]

[Fernandez et al., 1997] criticise the waterfall and the incremental lifecycles and propose to use the evolving prototype lifecycle that lets the ontologist modify, add, and remove definitions in the ontology at any time if some definition is missed or wrong. Knowledge acquisition, documentation and evaluation are support activities that are carried out throughout these states. When applied to a changing domain, the ontology will have to evolve. At anytime someone can ask for including or modifying notions of the ontology. To maintain the ontology is an important activity to be done carefully." [Gandon 2002 ch 2.1.4]

In all the different methodology of ontology design, the life cycle of an ontology development process is composed of several iterative steps:

1. Ontology specification
2. Knowledge acquisition
3. Conceptualization
4. Formalization
5. Evaluation
6. Documentation

3.1 The ontology specification

The ontology specification is needed to specify what is expected of the ontology. This specification determine : the purpose and the domain of the ontology.

3.1.1 the purpose of the ontology.

For example an formal / computational ontology about tourism can be used in a web engine in order to find appropriate answer about transport reservation or trip conception (transport+ location). A descriptive ontology can be used for teaching a specific domain and organize documentation and exercise for student. The intended users/reader of the ontology and what is their purpose has to be determined. This purpose can be set by listing typical query that the ontology has to answer or by describing a usage scenario.

3.1.2 The domain of the ontology.

The domain fix the set of objects/tasks described by the ontology and the point of view about it: Who is going to use this object/task for which purpose? The domain ontology is used to know if an arbitrary object is out of the scope of the ontology in order to define the ontology boundaries.

3.2 The knowledge acquisition

This stage begins by gather all available knowledge resources describing the domain of the ontology. This resources can be:

- ontologies : top level ontology, existing domain ontology
- glossary of terms: dictionaries, thesaurus
- technical documentation, image, report etc…
- tools, algorithms etc…

knowledge can be elicited using techniques such as: brainstorming, users observations, interviews, questionnaire/ questioning, formal and informal analysis of texts, and knowledge acquisition tools [Fernandez et al., 1997].

[Aussenac-Gilles et al., 2000] promote a new approach for knowledge modelling based on knowledge elicitation from technical documents which benefits from the increasing amount of available electronic texts and of the maturity of natural language processing tools. The acquisition of knowledge from texts applies natural language processing tools and based on results in linguistics to semi-automate systematic text analysis and ease the modelling process.

The result of this phase is to identify the most important terms of the domain and define them thanks to the more consensual definitions. At this stage a term is considered as a concept even if theoretically a term is only a linguistic representation of a concept in a text. Thus, the same term can have a different interpretation in a different text.

The term definition should indicate the semantic of the terms and his context that is who use this terms during which process. We have to find consensual definition so the best to do is to find definition sited in consensual book well known in the domain of study. A set of metadata can be added to specify the use context of the term for example the context can define which actors uses the term during which process.

"[Uschold and Gruninger, 1996] give guidelines to generate the definitions:

- *Produce a natural language text definition, being as precise as possible.*
- *Ensure consistency with terms already in use; in particular:*
- *Make ample use of dictionaries, thesauri and other technical glossaries.*
- *Avoid to introduce new terms where possible.*
- *Indicate the relationship with other commonly used terms that are similar to the one being defined (e.g. synonyms or variants with same underlying notion, but perhaps from different perspectives).*
- *Avoid circularity in defining terms; this increases clarity in general and is essential for later formalisation.*
- *The definition of each term is intended to be necessary and sufficient as far as this is possible in natural language. Provide clarification or additional information essential to understanding the definition as separate notes following the definition."* ,,[Gandon 2002]

3.2.1 Terminology normalisation

At the end of the knowledge acquisition phase a lexicon is build where each term is defined by the more consensual and precise definition. This lexicon should be normalize during the conceptualisation phase in order to regroup synonyms representing the same concept or disambiguate ambiguous terms denoting different concepts. A definition should not use a no defined terms. During normalisation, the amount of data to be studied is gradually restricted.

3.3 Conceptualization

This passage from linguistic wording to the formal notions is bridged by the rigorous lexicons that were obtained in the previous stage. The goal of this stage is to detect, define and organize concept. During this phase a concept is no more a term, but a concept is a definition. Metadata or attribute can be used to characterize concepts.

3.3.1 Concept detection

Thanks to the terminology build during the knowledge acquisition, concept should be detected. That means choose which terms are to be conceptualize. For example, the ontologist can choose to regroup terms no significantly different to build a single concept etc…

the conceptualisation process is a refinement activity, iteratively producing new refined representations of the notions being modelled (intermediary representation as they are called in [Gómez et al., 1996]). This is the beginning of a continuum between informal textual definitions of the notions at play in a scenario, and implemented representation of these notions in formal languages. Every new intermediary representation is a step toward the required degree of formalisation.

"[Uschold and Gruninger, 1996] experienced considerable variation in the degree of effort required to agree on definitions and terms for underlying concepts: "For some terms, consensus on the definition of a single concept was fairly easy. In other cases several terms seemed to correspond with one concept definition. In particular, there were several cases where commonly used terms had significantly different informal usage, but no useful different definitions could be agreed. This was recorded in notes against the definition. Finally, some highly ambiguous terms are identified as corresponding with several closely related, but different concepts. In this situation, the term itself gets in the way of a shared understanding."

Thus, there exist three cases when studying terms:

- *One term corresponding to one and only one definition: there is no problem.*
- *Several terms corresponding to one definition: The terms are synonyms, one should keep the list of synonyms and choose one preferred term to refer to that definition.*
- *One term corresponding several concepts: This corresponds to an ambiguity: homographs.*

When faced with an ambiguous term, [Uschold and Gruninger, 1996] suspend use of the term; they give the different notions meaningless labels such as x1, x2, x3 etc. so they can be conveniently referred to in a neutral way; they clarify the notions by carefully defining them using as few technical terms as possible, or only those whose meaning is agreed (consult the dictionary, thesauri, and/or other technical glossaries); they determine which, if any, of the concepts are important enough to be in the ontology; finally, they choose a wording for these notions ideally avoiding the original ambiguous ones." [Gandon 2002, ch 2.2.3]

[Uschold and Gruninger, 1996] structure the terms loosely into work areas corresponding to naturally arising sub-groups; groups arose such that terms are more related to other terms within the same group than they are to terms in other groups. This could be compared to clustering based on closest neighbours. They group similar terms and potential synonyms together into areas and then they identify semantic cross-references between the areas i.e., concepts that are likely to refer to or to be referred to by concepts in other areas. Authors suggest that this information can be used to help identify which work area to tackle first to minimise likelihood of rework. They address each work area in turn, starting with work areas that have the most semantic overlap with other work areas. These are the most important to get right in the first place, because mistakes lead to more re-work. [Uschold and Gruninger, 1996]

3.3.2 Relation detection

Conceptualising leads ontology designers to organise the notions and to do so they look for relations between these notions that could be used to structure them. The first relation, ontologist can use is the semantic similarity closed to the lexical fields. The second relation used is a partition, indeed, from the definitions, it rapidly emerges that some terms have different roles: some terms denote natural concepts and others denote relations between these concepts. The relations have their definition in which it is important to include the types of concepts they link. This is called the relation signature.

Relations can be labelled with verbs, or verbal groups, thus one criteria for partitioning linguistic expressions can be to separate verbs and non verbs; it is the case in the intermediary representations of [Fernandez et al., 1997].

[Maedche and Staab, 2000] describe an approach for discovering non-taxonomic conceptual relations from text and for facilitating this part of ontology engineering.

3.3.3 Concept taxonomie construction

Some researchers in IA are specialised in knowledge modelling and ontology design. Thus several methodologies have been proposed in order to build a taxonomy of concepts based on a hierarchical relation. This hierarchical relation is the subsomption relation (specialisation/generalisation relation) In particular in ontology design, several approaches have been opposed in literature [Gandon, 2002]:

- Bottom-Up approach proposed in [Van Der Vet, 1998]: starting from the most specific concepts and building the conceptual hierarchy by generalisation; the ontology is built by determining first the most specific concepts and by generalising them. This approach is prone to provide tailored and specific ontologies with fine detail grain concepts.
- Top-Down approach recommended by Sowa [Sowa, 1995]: Starting from the most generic concept and building a structure by specialisation; the ontology is built by determining first the top concepts and by specialising them. This approach is prone to the reuse of ontologies and inclusion of high level philosophical considerations which can be very interesting for coherence maintenance.
- Middle-Out approach of Uschold and Gruninger [Uschold et al., 1996]: Identifying core concepts in each domain identified and then generalised and specialised them to complete the ontology. This approach is prone to encourage emergence of thematic fields and to enhance modularity and stability of the result.

Others approach based on semantics have been proposed. These methods drive the construction on concepts taxonomy [Gandon, 2002] by explaining the structural principle.

“ [Bachimont, 2000] proposes to determine the meaning of a concept in the tree using four differential principles. When applied to a given concept, these principles make explicit similarities and differences with its neighbours. These principles are:

- *The parent community principle: any concept is determined by the similarities it has in common with its parent; This principle is at the core of the generalisation process.*
- *The parent difference principle: any concept is different from its parent, otherwise there would be no point to define it; one must make explicit the difference that distinguishes it from its parent concept. This principle is at the core of the specialisation process.*
- *The brother difference principle: any concept is different from its brothers, otherwise there would be no point to define it; one must make explicit the difference that distinguishes it from its brother concepts.*
- *The brother community principle: by definition all the children concepts of a parent concept have in common the same generic feature as the one they share with their parent concept. But one must establish another community between children concepts: the one that enables us to define the differences mutually exclusive between children concepts.*

Thus, Bachimont approach define only one type of difference with the parent concept. [Kassel et *al., 2000] propose to specialise a concept in several ways. For instance the concept 'packaging' could be specialised along several types of differences: "recyclable / non recyclable", "shock-proof / fragile", etc. "By considering the definitions we can bring out similarities between the differentia. (...) These groupings correspond to the notion of semantic axis" [Kassel et al., 2000]. A semantic axis groups the children of a concept according to the characteristics involved in the definition of their differentia. This notion reifies the brother community and difference principles. For instance: documents distinguished by their medium (paper, digital...), documents distinguished by their role (message, reference...), document distinguished by their status (official, informal, etc.), and so on.”* [Gandon 2002 ch 2.2.5.1]

“[*Guarino and Welty, 2000] also made a significant contribution to the theoretical foundations of the ontology field. In 1992, Guarino started by distinguishing natural concept, role, attributes, slots, qualities. He proposed to use the term 'role' only in Sowa's sense; bearing on Husserl's theory of foundation, he distinguishes between roles and natural concepts, and defines a role as a concept which implies some particular 'pattern of relationships', but does not necessarily act as a conceptual component of something. He defines 'attributes' as concepts having an associate relational interpretation, allowing them to act as conceptual components as well as concepts on their own; He proposes a formal semantics which binds these concepts to their corresponding relations, and a linguistic criterion to distinguish attributes from 'slots', i.e., from those relations which cannot be considered as conceptual components. Moreover, he shows how the choice of considering attributes as concepts enforces 'discipline' in conceptual analysis as well as 'uniformity' in knowledge representation. [Guarino, 1992].”*[Gandon 2002 ch 2.2.5.3]

The extensive work of Guarino contributes to clean-up the theoretical foundations of ontology engineering. [Guarino and Welty, 2002] propose 4 constraints definitions to be verified by the taxonomy:

- Rigidity constraints
- Identity constraints
- Unity constraints
- Dependence constraints

So that an ontologist relying on these definitions can check some validity aspects of the taxonomy.

The only problem so far is that no tool is available to help an ontologist do that work independently from a formalism, and it can become titanic to apply this theory to large ontologies. IODE from Ontology Works [Guarino and Welty, 2002] is an interesting integrated tool that is currently being used in several projects.

3.3.4 Relation taxonomie construction

In addition to terms corresponding to concepts, the ontologist also gathers terms corresponding to relations. [Bachimont, 2000] declares that the definitions of the relations must be based on the concepts they link. Considering binary relations the author defines them the following way:

- The semantic signature of the relation: the concepts linked up by the relation constitute its signature which is part of its definition.

- The intrinsic semantics of the relation: it is its intension, specified by comparison to other relations according to the differential paradigm. The signature is not sufficient: for example, the relations 'voluntary agent' and 'involuntary agent' besides to have in common their signature, also have in common the parent relation 'agent' while their intensions are different.

3.4 Formalization

The formalization goal is to build formal representation of an ontology. Thus ontologist has to choose a language to capture the content of the intermediary representation we built. The taxonomy is one of them, but the needed expressiveness may be higher.

The next stage is to populate the ontology to build the knowledge base. This stage add instance of concept in order to describe some situation that is the formal knowledge.

3.5 Evaluation

Different aspects of ontology can be evaluate:

- The satisfaction of users when testing the ontology. For example does the ontology help them to process more rapidly their task?
- The completeness of the domain representation: is there any ambiguous concept/relation, missing concept/relation or superfluous concept/relation ?
- The correctness of the knowledge base and its inference engine [Guarino and Welty, 2002]. Is there any cycle in the concept taxonomy? Is the concept place correctly in the taxonomy? Does the inference engine produce falls knowledge ?

Some metrics has to be defined in order to evaluate all this aspects: [Brewster et al 2004], [Gomez Perez 1999].

3.6 Documentation

During the whole process of ontology development each choice or problem has to be documented and explained like in project management. All the definition found has to be documented too in order to precise the source documentation and the authors.

4 Index

notion something formed in the mind, a constituent of thought; it is used to structure knowledge and perceptions of the world. || an idea, a principle, which can be semantically valued and communicated. [Gandon 2002]

concept notion usually expressed by a term (or more generally by a sign) || a concept represents a group of objects or beings sharing characteristics that enable us to recognise them as forming and belonging to this group. [Gandon 2002]

relation notion of an association or a link between concepts usually expressed by a term or a graphical convention (or more generally by a sign) [Gandon 2002]

extension / intension distinction between ways in which a notion may be regarded: its extension is the collection of things to which the notion applies; its intension is the set of features those things are presumed to have in common. There exists a duality between intension and extension: to included intensions $I1 \subset I2$ correspond included extensions $E1 \supset E2$. [Gandon 2002]

concept in intension / intension of a concept set of attributes, characteristics or properties shared by the object or beings included in or to which the concept applies. e.g. for the concept of a car the intension includes the characteristics of a road vehicle with an engine, usually four wheels and seating for between one and six people. [Gandon 2002]

concept in extension / extension of a concept set of objects or beings included in or to which the concept applies.e.g. for the concept of a car the extension includes: the Mazda MX5 with the registration 2561 SH 45, the green car parked at the corner of the road in front of my office, etc. [Gandon 2002]

relation in intension / intension of a relation set of attributes, characteristics or properties that characterises every realisation of a relation.e.g. for the relation parenthood the intension includes the characteristics of the raising of children and all the responsibilities and activities that are involved in it. [Gandon 2002]

signature of a relation set of concepts that can be linked by a relation, this constraint is a characteristic of the relation that participate to the definition of its intension.e.g. for the relation parenthood the signature says it is a relation between two members of the same species. [Gandon 2002]

relation in extension / extension of a relation set of effective realisations of a relation between object or beings.e.g. for the relation parenthood the extension includes: Jina and Toms are the Parents of Jim, Mr Michel Gandon is my Father, etc. [Gandon 2002]

Ontology that branch of philosophy which deals with the nature and the organisation of reality [Guarino and Giaretta, 1995]. || a branch of metaphysics which investigates the nature and essential properties and relations of all beings as such. [Gandon 2002]

Formal Ontology the systematic, formal, axiomatic development of the logic of all forms and modes of being [Guarino and Giaretta, 1995].

conceptualisation an intensional semantic structure which encodes the implicit rules constraining the structure of a piece of reality [Guarino and Giaretta, 1995] || the action of building such a structure.

ontology a logical theory which gives an explicit, partial account of a conceptualisation [Guarino and Giaretta, 1995] (based on [Gruber, 1993]); the aim of ontologies is to define which primitives, provided with their associated semantics, are necessary for knowledge representation in a given context. [Bachimont, 2000]

ontological commitment a partial semantic account of the intended conceptualisation of a logical theory [Guarino and Giaretta, 1995] || practically, an agreement to use a vocabulary (i.e., ask queries and make assertions) in a way that is consistent with respect to the theory that specifies the ontology. Software pieces are built so that they commit to ontologies and ontologies are designed so that they enable us to share knowledge with and among these software pieces. [Uschold and Gruninger, 1996]

ontological theory a set of formulas intended to be always true according to a certain conceptualisation [Guarino and Giaretta, 1995].

ontological engineering the branch of knowledge engineering which exploits the principles of (formal) Ontology to build ontologies [Guarino and Giaretta, 1995]. || defining an ontology is a modelling task based on the linguistic expression of knowledge. [Bachimont, 2000]

ontologist a person who builds ontologies or whose job is connected with ontologies' science or engineering. [Gandon 2002]

state of affairs the general state of things, the combination of circumstances at a given time. The ontology can provide the conceptual vocabulary to describe a state of affairs. Together this description and the state of affair form a model. [Gandon 2002]

taxonomy a classification based on similarities. [Gandon 2002]

term: term is the lexical representation of a concept

Term : linguistique representation of a concept in a text. « manifestation linguistique d'un concept dans un texte. En terminologie, le terme est une désignation au moyen d'unité linguistique d'une notion définie en langue de spécialité » (vocabulaire de la terminologie : norme AFNOR X03-003).

Mereology The study of part-whole relationships. [Gandon 2002]

partonomy a classification based on part-of relation. [Gandon 2002]

Table 2 Definitions used in ontology engineering based on [Gandon 2002]

5 Bibliographie

[Aussenac-Gilles et al., 2000] Aussenac-Gilles N., Biebow B, Szulman S (2000). Revisiting Ontology Design: a Method Based on Corpus Analysis, In Proc. EKAW'2000, Juan-les-Pins, p172-188.

[Bachimont, 2000] Bruno Bachimont: "Engagement sémantique et engagement ontologique: conception et réalisation d'ontologies en ingénierie des connaissances"; In "Ingénierie des connaissances Evolutions récentes et nouveaux défis", Jean Charlet, Manuel Zacklad, Gilles Kassel, Didier Bourigault; Eyrolles 2000, ISBN 2-212-09110-9

[Brewster et al 2004] Brewster, Christopher and Alani, Harith and Dasmahapatra, Srinandan and Wilks, Yorick (2004) Data Driven Ontology Evaluation. International Conference on Language Resources and Evaluation, Lisbon, Portugal.

[Gandon, 2002] Fabien Gandon "Distributed Artificial Intelligence And Knowledge Management: Ontologies And Multi-Agent Systems For A Corporate Semantic Web ", scientific philosopher doctorate thesis in informatics INRIA and University of Nice - Sophia Antipolis — Doctoral School of Sciences and Technologies of Information and Communication (S.T.I.C.), 7/11/2002, 486 pages

[Gomez Perez 1999] Asuncion Gomez-Perez, 'Evaluation of taxonomic knowledge in on- tologies and knowledge bases', in Proceedings of the 12th Bannf Knowledge Acquisition for Knowledge-based Systems Workshop, Banff, Canada, (1999).

[Gruber 1993]T. R. Gruber. A translation approach to portable ontologies. Knowledge Acquisition, 5(2):199-220, 1993

[Gruber web site] http://www-ksl.stanford.edu/kst/what-is-an-ontology.html

[Guarino, 1992] Guarino N. Concepts, Attributes, and Arbitrary Relations: Some Linguistic and Ontological Criteria for Structuring Knowledge Bases. Data and Knowledge Engineering 8: 249-261, 1992.

[Guarino and Welty, 2000] Guarino, N. and Welty, C. 2000. Towards a methodology for ontology-based model engineering. In Proceedings of ECOOP-2000 Workshop on Model Engineering. Cannes, France. Available from http://www.ladseb.pd.cnr.it/infor/ontology/Papers/OntologyPapers.html

[Guarino and Welty, 2002] Guarino, N. and Welty, C. 2002. Evaluating Ontological Decisions with OntoClean. Communications of the ACM, 45(2): 61-65.

[Kassel et al., 2000] The project is presented in "Construction et exploitation d'une ontologie pour la gestion des connaissances d'une équipe de recherche" Kassel G., Abel M.-H., Barry C., Boulitreau P., Irastorza C., Perpette S.; Proceedings of IC'2000, Toulouse, 10-12 Mai 2000, available at http://www.irit.fr/IC2000/actes-enligne.html

[Maedche and Staab, 2001] Maedche A. and Staab S., Comparing Ontologies — Similarity Measures and a Comparison Study, Internal Report No. 408, Institute AIFB, University of Karlsruhe, 76128 Karlsruhe, Germany, March 2001

[Studer 1998] Rudi Studer, V. Richard Benjamins, Dieter Fensel: Knowledge Engineering: Principles and Methods. Data Knowl. Eng. 25(1-2): 161-197 (1998) [Sowa, 2000] Guided Tour of Ontology; John F. Sowa http://www.jfsowa.com/ontology/guided.htm

[Sowa, 1995] Sowa, J. "Top-Level Ontological Categories". In International Journal on Human-Computer Studies, 1995, Vol. 43, N° 5/6, pp. 669-685.

[Uschold and Gruninger, 1996] M. Uschold and Gruninger M. Ontologies: Principles, methods and applications. Knowledge Engineering Review, Vol. 11:2, 93-136, 1996. Also available as AIAI-TR-191 from AIAI, The University of Edinburgh.

[Van der Vet, 1998] Van Der Vet, P.E., Mars, N.J.I. "Bottom-up construction of ontologies". In IEEE Transactions on Knowledge and Data Engineering, June-August 1998, Vol 10, N°4. pp. 513-526.