UNIVERSITY
OF TRENTO

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

38050 Povo – Trento (Italy), Via Sommarive 14
http://www.dit.unitn.it

DERIVING ONTOLOGY-BASED METADATA FOR E-LEARNING FROM THE ACM COMPUTING CURRICULA

Paramjeet Saini and Marco Ronchetti

May 2003

Technical Report # DIT-03-017

DERIVING ONTOLOGY-BASED METADATA FOR E-LEARNING FROM THE ACM COMPUTING CURRICULA[1]

Paramjeet Saini
Dipartimento di Informatica e Telecomunicazioni
Università degli Studi di Trento, 38050 Povo (Trento) Italy
pssaini@science.unitn.it

Marco Ronchetti
Dipartim ento di Informatica e Telecomunicazioni
Università degli Studi di Trento, 38050 Povo (Trento) Italy
Marco.Ronchetti@unitn.it

ABSTRACT

Using metadata is very important for achieving best results in finding and reusing elearning material. Metadata are however effective only if there is a common agreement on the terminology used. Defining an ontology is a way out, but creating a good ontology and reaching a consensus on it is a non trivial task. In this paper, we claim that an ontology for Computer Science can be extracted as a side-result from the work made by an ACM committee on Computer Science curricula. We extracted the ontology, and we suggest that it could/should be used for defining metadata on elearning material that is concerned with the Computer Science domain. We provide both an XML and a DAML+OIL representation of the ontology. Also, we suggest a possible way to navigate the metadata that allows enriching each lesson with a set of related material automatically extracted from the repository of e-learning material. Finally, we discuss possible additional uses of the ontology, and the exportability of our approach to other disciplines.

KEYWORDS

Metadata, Ontology

1. INTRODUCTION

It is well known that the production of elearning material is an expensive task: therefore the perspective of being able to "write once, use anywhere", i.e. the ability to reuse learning material is a very appealing one. Unfortunately, as software engineers know, writing for reuse is difficult and costly, and an infrastructure is needed to be able to effectively find reusable material. Metadata's mission is to facilitate reuse in different environment. Some metadata are "container oriented", i.e. they are meant to inform the learning management system about format, sequencing etc. of the material. Other metadata are meant to facilitate the use of the material by teachers, by curriculum designers and by producers. Several standards have come out to define a common convention for metadata definition and representation (IEEE LTSC P1484.12, , AICC AGR, ADL SCORM, Dublin CORE METADATA INITIATIVE etc.).

While container oriented metadata are well accepted and used, there are ongoing discussions on the other kinds of metadata that create conflicts and hamper the adoption of meta-data technologies. An interesting point of view is expressed by [Nilsson 2002]: they argue that the image of meta-data as being objective information about data is wrong, or at least incomplete. This image is tied to the fact that most meta-data aware systems only contain indisputable information such as title, author, identifier, etc. However, data description about the type of granularity of objects, pedagogical purpose, assessments and learning objectives, etc., represent subjective interpretations of resources. They therefore support the existence of

[1] Paper accepted at the IADIS e-Society Conference, Lisboa 3-6 June 2003

multiple, even conflicting descriptions, and maintain that the RDF technology allow to implement and use such subjective metadata.

However, even to express subjective views we need to agree on a common vocabulary that specifies what we mean. That's where the notion of ontology comes in. According to Gruber "A specification of a representational vocabulary for a shared domain of discourse -- definitions of classes, relations, functions, and other objects -- is called an ontology" [Gruber 1993]. An ontology is therefore a controlled, hierarchical vocabulary for describing a knowledge system. It abstracts the essence of concepts, and allows distinguishing various kinds of objects and defining the relationships among them.

Besides the subjective vs. objective metadata issue, we believe that the part of the problem with metadata and with reusability stems from the lack of a common ontology. In the present paper we propose an ontology for describing the content of elearning material in a particular domain: Computer Science. We also discuss the extensibility of our model to other disciplines.

2. DEFINING AN ONTOLOGY

The advantages that an ontology offer are obvious: "Gradually, computer scientists are beginning to recognize that the provision, once and for all, of a common, robust reference ontology – a shared taxonomy of entities – might provide significant advantages over the ad hoc, case-by-case methods previously used." [Smith 2001]. The problem is: how is it possible to produce an ontology?

A good ontology should cover the target domain in an exhaustive way. Moreover, an essential property is that it should be accepted by a broad community: ontological commitment is defined as the agreement of multiple parties to adopt a particular ontology when communicating about a specific domain. Both these features are very difficult to achieve. Therefore some research has been devoted to solving the problem of how to construct a good ontology. [Holsapple 2002] list five possible approaches to ontology design: Inspiration, Induction, Deduction, Synthesis and Collaboration. Inspiration is based on individual viewpoint about the domain; Induction and Deduction start respectively from specific cases and from general principles relative to the domain. Synthesis puts together existing ontologies to generate more general or more agreed-upon ones. Collaboration is somehow similar to synthesis, but while synthesis puts together several efforts by combining the finite products, collaboration obtain the result by putting together the processes that would generate individual ontologies. According to [Holsapple 2002], the last approach has a built-in evaluation facility to access quality and acceptability of the resulting ontology. It their conclusion, they claim that a collaborative approach can be used to design ontologies for many applications, among which distance learning.

We state that in some fortunate cases yet another possible approach is possible, i.e. to reuse (possibly implicit) ontologies that were agreed upon within a framework in a different setting. We demonstrate such case. In 2001 the Association for Computing Machinery (ACM) published the result of an excellent work for recommending undergraduate program in computer science. We notice that in doing that, they went through the definition of an ontology that has both the nice properties above mentioned: it covers exhaustively a domain, and is the product of a large, collective work that encountered a very broad acceptance in the Computer Science community. We therefore propose to reuse such ontology for expressing the computer-science-domain-specific metadata of e-learning artifacts.

3. THE ACM COMPUTING CURRICULA 2001 FOR COMPUTER SCIENCE

Since forty years, ACM gives recommendations for the undergraduate program in computer science: the process produces a new recommendation approximately every ten years [ACM1965,ACM1968,ACM1978, Tucker 1991] Recently, a new recommendation was released in its final form [ACM2001]: the Computing Curricula 2001 for Computer Science (CC2001). More than 150 people were directly involved in the focus groups established to contribute to the process. Their work took three years: the task was defined in 1998, and the final version is dated December 2001. The stated goal was "to review the Joint ACM and IEEE/CS Computing Curricula 1991 and develop a revised and enhanced version for the year 2001 that will match the latest developments of computing technologies in the past decade and endure through the next decade". To ensure the broad participation necessary for success of the project, fourteen knowledge focus groups representing a wide range of constituencies and areas of expertise were established. Six pedagogy focus groups were in charge of developing a holistic perspective and to address a variety of questions that transcend the boundaries of the individual subdisciplines. The report has been widely reviewed by academics and practitioners through a series of three public drafts, and feedback was obtained through sessions at conferences and meetings, including the Special Interest Group on Computer Science Education symposium (SIGCSE), the Frontiers in Education conference (FIE), the World Congress on Computers and Education (WCCE). The final document was endorsed by the ACM Council in November 2001 and by the IEEE-CS Board of Governors in December 2001.

Although the definition of CC2001 is strongly influenced by educational practice in the United States, the intent was to ensure that the curriculum recommendations are sensitive to national and cultural differences so that CC2001 may be useful to computing educators throughout the world (for instance, at the University of Trento, Italy, we found it to be very useful for checking content and structure of our own CS curriculum). So we can conclude that the work done by ACM meets the requirements to cover a domain in an exhaustive way, and to be accepted by a very broad community. However, its goal was not to define an ontology usable in the e-learning domain, but rather to be a reference for building a traditional college curriculum.

We notice that in the process of defining the curricula, and intermediate step is the definition of an ontology. We propose that this intermediate step be used as the ontological foundation for building metadata that allow indexing and reusing e-learning material in the Computer Science domain. Moreover, it can be used to provide a metalevel navigation tool that enhances the usability of elearning material. To illustrate the reusability of CC2001 as an ontological foundation, we shortly describe the structure of the results that it provides.

One final result is the definition of a suite of courses, subdivided into three categories: introductory, intermediate and advanced. Prerequisites and syllabus are specified for each course (actually, only the 47 introductory and intermediate course are covered in detail: the 80 advanced courses are not fully described in the document). Another final product is the set of "Curriculum models". The report identifies six approaches to introductory computer science that have proven successful in practice, four thematic lines for presenting the intermediate-level courses, and some other example of curricula as a whole.

As we mentioned however, the most important part for our goals is the intermediate step that allows defining courses and curricula: the definition of the Computer Science body of knowledge, i.e. the catalog of "knowledge elements" appropriate to undergraduate computer science programs. The CS body of knowledge is organized hierarchically into three levels. The highest level of the hierarchy is the area, which represents a particular disciplinary subfield (e.g. "Human-Computer Interaction"). The areas are broken down into smaller divisions called units, which represent individual thematic modules within an area. Each unit is further subdivided into a set of topics, which are the lowest level of the hierarchy. For instance, the topic "Online communities: MUDs/MOOs" belongs to the unit "HCI aspects of collaboration and communication" that is part of the area "Human-Computer Interaction". Each unit has a number of learning objectives that are associated to it: for instance, one of the learning objectives of the unit "HCI aspects of collaboration and communication" is to "discuss several issues of social concern raised by collaborative software" (typically there are approx. 5 learning objectives per unit).

In total, the body of knowledge is divided in 14 areas, 132 units and 950 topics. The units are further classified as belonging to the "core" or being "elective", the core being composed by the 64 units for which there is a broad consensus that the material is essential to an undergraduate degree in computer science. We believe that the expression of body of knowledge is a perfect ontology for describing e-learning material in the field of Computer Science: it is granular enough to be precise, it has aggregation relations that allow finding related/similar material, and defines a vocabulary that removes ambiguity.

4. REPRESENTATIONS OF THE ONTOLOGY

The CC2001 document comes with two appendixes: Appendix A-CS Body of Knowledge and Appendix B-Course Description.

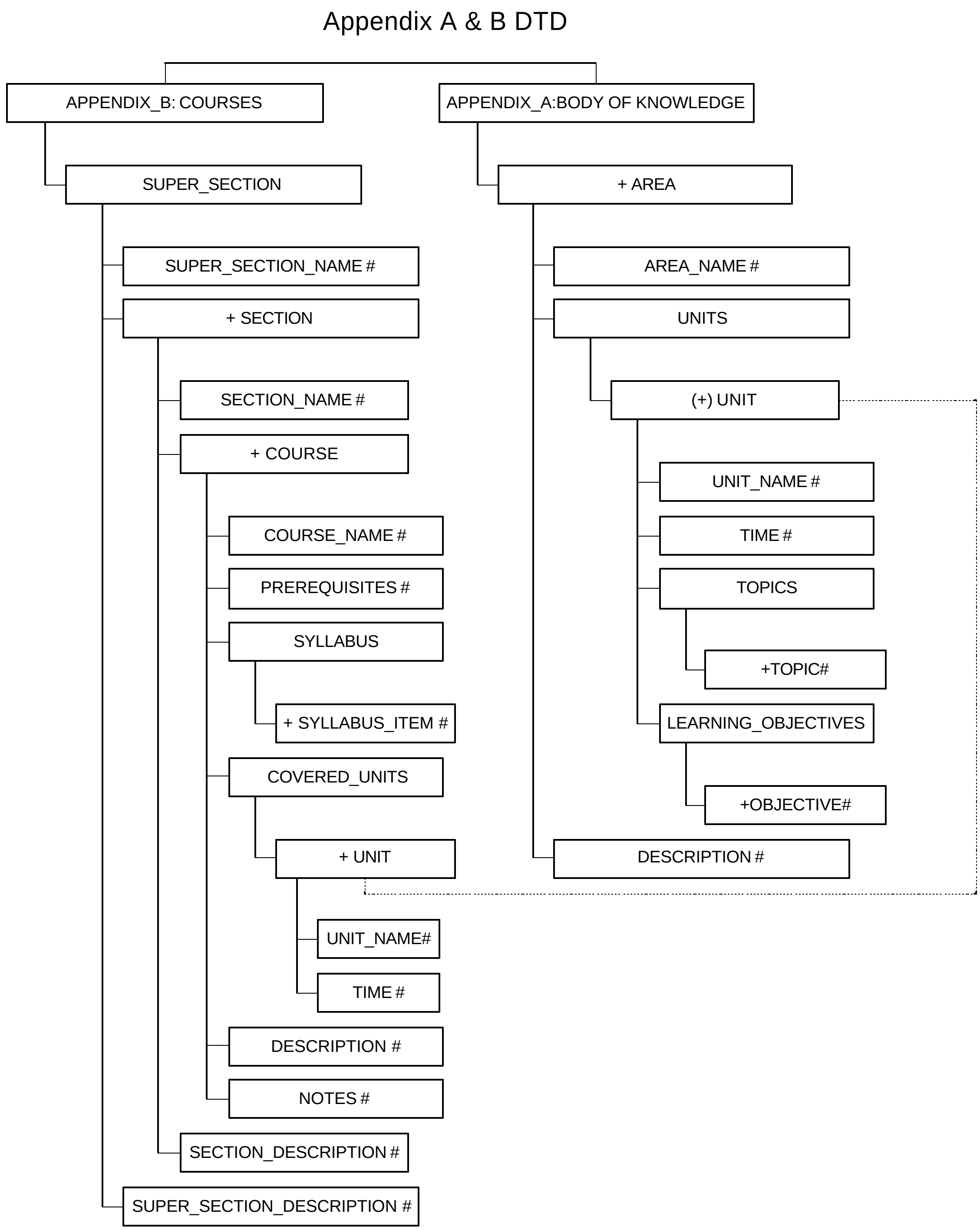

Figure 1 – The DTD structure for the XML description of the Computing Curricula 2001 Computer Science. + indicates "one or more instances", # indicates that the node is a leaf (although it might contain CDATA). The UNIT tag is present in both Appendixes, and referrers to common data.

Appendix A contains the detailed structure of the Body of knowledge: for each area is given a short description and the list of units. For each unit is specified if the unit belongs to "core", a measure of its extent (according to a metric measured in conventional units called "hours", although the document does not endorse a particular teaching style), a list of topics, and a list of learning objectives.

Appendix B contains a list of courses, grouped in sections (introductory, intermediate and advanced). Sections may be divided in alternative tracks (object-first, hardware-first etc.).

Courses contain a description, a list of prerequisites, syllabus, a list of covered units and sometimes additional notes. ("Units" are basic building blocks and are referred in both Appendices).

We extracted information from the CC2001 pdf file by defining a grammar that could feed a parser to automatically extract data and format them in the from of an XML file. We could not define a grammar that would respond to all the needs: although the Appendices are highly structured, their structure was probably not formally defined. Moreover, accidents like page numbers ended up in the text file that was passed to the parser. We had therefore to do some fine-tuning by manually adjusting the XML data. Finally, a DTD was defined for each Appendix so that we could check the syntactic correctness of the resulting files.
Figure 1 shows a graphic representation of the structure defined by the DTD.

The representation provided by the DTD+XML files is "pure" in the sense that it reflects the structure given by the ACM committee. However it is not completely satisfactory because it is limited to the representation of a taxonomy. An ontology could add more semantic relations than the simple containment. The simple XML representation however is not suited to incorporate additional relations, such as "is needed for", "is important for" that would be useful to select related material according to its relevance. These and similar relations would increase the usefulness of the ontology, and therefore it is important to choose a representation that allows their inclusion. We therefore transformed the primitive XML data in a DAML+OIL representation [Horrocks 2002]. Of course, by extending the original data with added information we risk loosing the wide consensus that should be implicit in the original representation.

XML and DTD files for Appendix A and Appendix B, as well as the DAML+OIL representation of the ontology are available on the web site http://latemar.science.unitn.it/Ontology .

3. POSSIBLE USES OF THE ONTOLOGY

As we mentioned, defining a common vocabulary and having good metadata enhances the chance of finding material (while studying, or while producing new lessons) and makes reusability possible. It also allows better knowledge management, information exchange and enables intelligent agents.

Also, it allows different navigation paths that pass (implicitly or explicitly) through metadata space. We are currently exploring such navigation. We envision a navigation system that allows students to perform a kind of navigation that is traditionally not allowed by standard hypertextual systems. In particular, we wish to allow navigation on a metalevel: at any time a student should be able leave the e-lecture s/he is currently engaged in, and move to a more abstract level (i.e. the metadata level), finding hyperlinks to related topics, and then plunging down in material that deals with the chosen topic. The idea is expressed in Figure 2.

In the figure, the lower plane represents the traditional hypertextual space (we will call that "content space"). Elements within the plan represent learning material, and arrows in the plane symbolize hyperlinks. The upper plane is the ontology plane. Entities in this plane are defined by ontology vocabulary. Thin arrows in the ontology plane stand for the relations among entities. A student's path (shown as the set of thick arrows) could start from the learning material, jump on the metadata level (i.e. on the ontology plane), follow relations on that plane and go back to learning material, performing an hyper-jump that was not anticipated by the producer(s) of the learning material: in fact it could end up looking at some learning object that belongs to a different course, but that is related to the material that was at the origin of the hyper-jump.

Such navigation could be started in many ways. An explicit form could be based on hyperlinks (to be placed for instance among the navigational options, typically as headers or footers of the page) that point to the "ontological space", When followed, these links would bring either to automatically synthesized pages that show the available links both in the ontological space and back to content space. Alternatively, the hyperlinks could activate a concept navigation tools, like the ones provided for navigating topic. Leaves in the topic maps should then expand in references to material in content space.

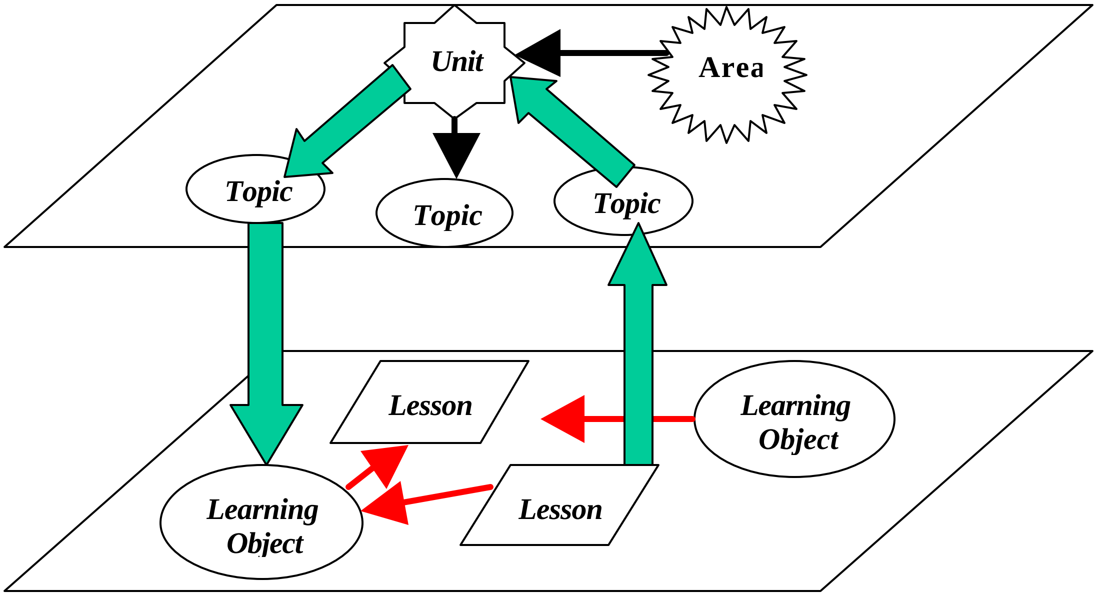

Figure 2: Traveling through content space and ontological space

We expect the explicit metadata navigation to be too abstract for students, but very useful for teachers who try to reuse learning material and to explicitly reference existing material (e.g. as a mean to cover prerequisites, or as additional material).

For students we are working at a different reification of the hyper-jumps idea. Each elearning document will be enriched by navigation headers and footers that point to "related material" and that opens an automatically generated page that list possible destinations points, grouped by (related) topic. The related topics can be chosen as the siblings of the topic to which the currently explored material belongs. In this way jumps through ontological space would happen without an active participation or awareness of the student. The tool that automatically compiles a list of related material and adds it at the end of each lecture in the repository is in the final development stage and should be ready by the time of the conference. We implemented such tool that will be described elsewhere.

4. EXPORTABILITY OF OUR APPROACH TO OTHER DISCIPLINES

A question is whether the approach we described can be applied to other disciplines. We are not aware of similar effort in other fields, although we believe that similar work could certainly be done for scientific disciplines: e.g. in Mathematics and Physics it should not be too difficult to produce similar results. Of course, the key of the Computer Science case is the wide recognition of the CC2001 model, that was officially endorsed by the major US association of the field (ACM and IEEE). If similar initiatives were started for instance by authoritative associations like AMS (American Mathematical Society) and APS (American Physical Society), a similar patterns could be applied to these disciplines. We have the feeling (but we would be happy to be proven wrong) that covering human science in a systematic, exhaustive and agreed upon way would be much more difficult.

Similar approaches could be thought by starting from different premises. The source we propose for the ontology is especially suited for providing e-learning meta-data because, after all, the original effort was anyhow concerned with teaching (although the focus was different). One could try to derive an ontology from different efforts, like the bibliographic codes (e.g. the Library of Congress classification), or the classification of scientific papers (like the one performed in the "Physics Abstracts"). However, we believe that both classifications have the wrong granularity: LOC classification and similar ones are too coarse, while the taxonomy of scientific papers is probably too fine. This second one is also perhaps out of focus for the goals we are interested in. So the question of how to apply our approach to other disciplines remains without a convincing answer.

7. CONCLUSION

We have claimed that ontologies are essential to have a language to express metadata in an e-learning environment. Building ontologies however is a very difficult task, in part because in order to reach the goal it is necessary to build consensus on the ontology itself. While some authors classify several ways to produce ontologies, we suggest that the most efficient way is to reuse existing ones. We have shown that an excellent ontology for the domain of Computer Science can be extracted as a byproduct from the ACM effort to suggest standard college curricula. We have explicitly extracted an XML and a DAML+OIL representation of the ontology. Moreover, we suggests that metadata built according to such ontological representation can be (explicitly or implicitly) navigated by teachers and students, leading to a more extensive and integrated use of the existing material .

We have discussed the possibility to extend our approach to other disciplines, concluding that there might be hope that a similar operation can be performed for other scientific discipline.

The approach we present has been used in practice: at Trento University we are presently in the process of putting on line e-learning material for many courses, and in particular we intend to publish on-line material for the whole set of Computer Science courses. For these courses, we label all the data with metadata derived by the ontology we presented here. Metadata navigation is made possible by suitable tools.

REFERENCES

ACM 1965 ACM Curriculum Committee on Computer Science. An undergraduate program in computer science—preliminary recommendations. *Communications of the ACM*, 8(9):543-552, September 1965.

ACM 1968 ACM Curriculum Committee on Computer Science. 1968 Curriculum '68: Recommendations for the undergraduate program in computer science. *Communications of the ACM*, 11(3):151-197, March 1968.

ACM 1978 ACM Curriculum Committee on Computer Science. 1978 Curriculum '78: Recommendations for the undergraduate program in computer science. *Communications of the ACM*, 22(3):147-166, March 1979.

ACM 2001 Computing Curricula 2001 Computer Science, *ACM Journal of Educational Resources in Computing, 1*(3):1-240, Fall 2001.

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

Holsapple C.W: and Joshi K.D. 2002 A collaborative approach to Ontology design, *Communications of the ACM*, 45(2):42-47, February 2002.

Horrocks. I. 2002 DAML+***OIL****: A description logic for the semantic web*. IEEE Bulletin of the Technical Committee on Data Engineering, 25(1):4–9, 2002.

Nilsson, Mikael, Palmér Matthias and Naeve Ambjörn, 2002 Semantic Web Metadata for e-Learning - Some Architectural Guidelines, WWW2002 International Conference

Smith, Barry and Chris Welty. 2001 *in Formal Ontology in Information Systems* "Ontology: Towards a new synthesis". In Chris Welty and Barry Smith, eds., Pp. iii-x. (Ongunquit, Maine, ACM Press 2001).

Tucker A.B. 1991 et al., Computing Curricula '91. Association for Computing Machinery and the Computer Society of the Institute of Electrical and Electronics Engineers, 1991.