MARI CARMEN SUÁREZ-FIGUEROA AND ASUNCIÓN GÓMEZ-PÉREZ

Abstract. In this paper we present the consensus reaching process followed within the NeOn consortium for the identification and definition of the activities involved in the ontology network development process. This work was conceived due to the lack of standardization in the Ontology Engineering terminology, which clearly contrasts with the Software Engineering field that boasts the *IEEE Standard Glossary of Software Engineering Terminology*. The paper also includes the NeOn Glossary of Activities, which is the result of the consensus reaching process here explained. Our future aim is to standardize the NeOn Glossary of Activities.

1. Introduction

The Semantic Web of the future will be characterized by using a very large number of ontologies embedded in ontology networks built by distributed teams (NeOn Consortium, 2006). So, future Semantic Web applications will be based on networks of contextualized ontologies, which will be in continuous evolution. With this new vision of the ontologies and the Semantic Web, it is important to provide strong methodological support for collaborative and context-sensitive development of networks of ontologies in distributed environments. This methodological support, which is being created within the NeOn project¹, includes the identification and definition of which activities should be carried out when ontology networks are collaboratively built.

The identification and definition of the activities involved in the development of ontology networks should be an agreed result within the Ontology Engineering field. However, it was noticed that researchers, technology developers, and users used different terminology to name the activities involved in the ontology development process; that is, no consensus

had been reached yet on many of the definitions for ontology engineering activities. For instance, it was not clear enough the difference between ontology modification (Stojanovic, 2004) and ontology update (Stojanovic et al., 2002); and other activities had multiple definitions in natural language (e.g., ontology merging (Fernández-López et al., 1997; Kalfoglou and Schorlemmer, 2003; Kotis and Vouros, 2004)). Additionally, new activities related to the Semantic Web of the future are emerging without a concrete and precise definition (e.g. ontology modularization). This terminological situation is the result of a lack of a standard in the Ontology Engineering terminology, which clearly contrasts with the Software Engineering field that boasts the *IEEE Standard Glossary of Software Engineering Terminology* (IEEE, 1990), which is a consensual glossary in the Software Engineering field.

Within the NeOn project, it has been decided to build the *NeOn Glossary* of *Activities* for unifying the terminology used by the NeOn partners. The goal was to achieve consensus on the identification and definition of the activities involved in developing ontology networks, which is the first attempt in the Ontology Engineering field towards to obtain a terminological standard.

To reach a consensus on the labels for naming the activities and on their definitions, we decided to use the wiki technology (Leuf and Cunnigham, 2001), which supports a high level of consensus building by community members because a user who disagrees with a statement (in our case labels and definitions) can very easily modify it, delete it, comment it, etc. (Viégas et al., 2004). We created a non public space in the NeOn wiki for discussing the Ontology Engineering terminology, expressing and exchanging different opinions among different partners involved in NeOn, and reaching a final agreement. Meetings and mailing lists were also employed for agreeing on activity definitions at final stages.

Thus, in this work we present the first step towards the standardization of the terminology used in the Ontology Engineering field. Such step lies in achieving consensus on the definitions for the activities involved in the development process for ontology networks, within the NeOn consortium, and building the NeOn Glossary of Activities. From our understanding, any standardization agency such as ISO or W3C deals with the unification of Ontology Engineering terms. Only some ISO technical committees such as ISO/TC37/SC4 are working on the contribution of the ontologies for unifying linguistic resources.

The rest of the paper is organized as follows: Section 2 presents the principles for definitions; Section 3 explains the consensus reaching process for the NeOn Glossary of Activities; Section 4 includes the NeOn Glossary of Activities; and finally, we include some conclusions in Section 5.

2. Principles for Defining

For collecting definitions during the consensus reaching process for the NeOn Glossary of Activities, we based on three basic principles of **defining** (Landau, 1984): "avoid circularity, define every word in a definition, and make sure that every word's definition says what the word means". In addition, we taken into account the following principles:

- Conciseness, i.e. every definition should say the most in the least number of words.
- Clarity in that it avoids ambiguity, i.e. words should be used unambiguously.
- Appropriateness, i.e. the definition should be appropriate to the target reader.
- Priority of essential traits, i.e. a definition should highlight the essential features of meaning.

3. Consensus Reaching Process for the Identification and Definition of the Activities

In this section we sketch the roles and the overall process gone through by the NeOn consortium to reach a consensus on the activities for developing ontology networks. During the process we have tried to achieve a consensus on the list of activities and on the activity definitions.

3.1. Roles in the Process

A varied number of skilled people, geographically dispersed (called 'NeOn Glossary' team), participated collaboratively in the consensus reaching process. The 'NeOn Glossary' team had a well-balanced and representative participation of people with different background: ontology engineers, ontology editors, and users within the NeOn project. The 'NeOn Glossary' team members were composed of 25 people belonging to 9 institutions. The following concrete roles were distinguished:

• *The NeOn Glossary coordinator*. One of the 9 institutions was in charge of creating the NeOn wiki page dedicated to the consensus reaching process. This institution created and included in the wiki a

template for gathering information about the activities identified in the literature for building ontologies. Additionally, the coordinator introduced for the debate the initial list of activities with initial definitions based on the study of the state of the art in the Ontology Engineering field.

• *The NeOn Glossary team*. The 9 institutions introduced definitions and commented them. They also participated in the different ad-hoc meetings carried out to reach consensus and used mailing lists for minor agreements.

3.2. Process Stages

Before beginning with the consensus reaching process, the meanings of consensus and consensus reaching process were explained. The proposed process of achieving consensus on the activities was also explained and reviewed by the 'NeOn Glossary' team. After that, the team agreed on a targeted time period (one year) to reach a consensus. And finally, the team followed the general process, shown in Figure 1, to achieve consensus on the activities involved in the development of ontology networks and on their definitions. We achieved consensus after the third round of the process.

The general process followed to achieve consensus on the activity terminology, shown in Figure 1, can be summarized as follows:

- 1. To create a NeOn wiki page dedicated to the consensus reaching process, within the NeOn consortium. We used this wiki page within the NeOn consortium to build collaboratively and by consensus the NeOn Glossary of Activities. This technology was chosen because wiki web sites facilitate to reach a consensus by community members (Viégas et al., 2004).
- 2. To create a template for gathering general information about activities and to publish this template in the wiki. The template includes the following slots: activity name, definition, type of activity according to IEEE, input, output, references and general comments.
- 3. To create an initial NeOn Glossary of Activities, including initial identification and definition of the main activities to be included in the ontology network development process. For creating this initial glossary, the NeOn Glossary coordinator manually extracted terms (in our case, activities for building ontologies) and their definitions from books, papers and documents well-known in the Ontology Engineering field. This initial glossary was made available on the wiki following the template created in step 2.

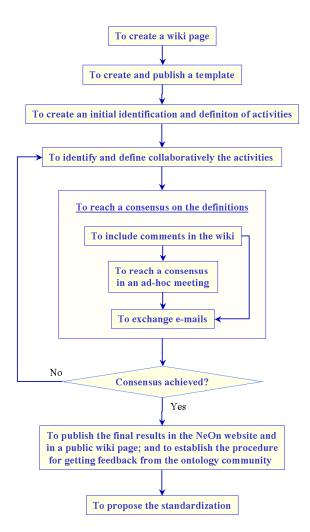


Figure 1 Consensus Reaching Process

- 4. To identify and define collaboratively the activities in the wiki, according to the initial list. People in the 'NeOn Glossary' team were totally free to incorporate more activities and/or definitions in the initial glossary; and to include more general information about the activity (such as, input, output, etc.).
- 5. To reach a consensus on the activity definitions. In this stage we used the wiki, ad-hoc meetings and mailing lists and we adopted the following process:
 - a. To include comments in the wiki. The 'NeOn Glossary' team introduced their comments for each activity in the wiki.
 - b. To reach a consensus in ad-hoc meetings. Two ad-hoc meetings and various e-mails were needed for reviewing and agreeing on the definitions of activities. The rules for accepting or not a

- particular activity definition are described in (Suárez-Figueroa and Gómez-Pérez, 2008).
- c. To exchange e-mails. Several e-mails were sent to the NeOn consortium to encourage other members not directly involved in the glossary creation to put forward their ideas.
- 6. To publish the final results in the NeOn website² and in a public wiki page and to establish the procedure for getting feedback from the Ontology Engineering community, using the argumentation tool Cicero³ (which is already a wiki).
- 7. To propose the standardization of the NeOn Glossary of Activities. The idea is to propose to standardization committees, such as the technical committee ISO/TC37, the standardization of the NeOn Glossary. Terminology standards help to avoid confusion by harmonizing terms, in our case activities involved in the development of ontology networks. The future standard NeOn Glossary of Activities is intended to serve as useful reference for those in the Ontology Engineering field and for those who come into contact with ontologies.

4. The NeOn Glossary of Activities

According to ISO 12200:1999 (ISO, 1999) and ISO 1087-1:2000 (ISO, 2000), a **glossary** can be defined as:

- a. A terminological dictionary containing the terminology of a specific subject field or of related fields and based on terminology work (ISO, 1999).
- b. A terminological dictionary that contains designations and definitions from one or more specific subject fields. The vocabulary may be monolingual, bilingual or multilingual (ISO, 2000).

Following the aforementioned definitions, we have collected in the *Neon Glossary of Activities* 53 activities involved in the ontology network development process (in the specific field of Ontology Engineering) and provided agreed natural language definitions and explanations of them. The vocabulary included in the glossary is monolingual (English). The list of activities included in the glossary could be not exhaustive, and for this reason we are yet working on the glossary for obtaining a complete list of activities.

^{2.} http://www.neon-project.org/

^{3.} http://cicero.uni-koblenz.de

As a result of the consensus reaching process, 51 activities were identified and defined as part of the NeOn Glossary of Activities, by means of the consensus reaching process explained in section 3. Apart from the 51 activities resulting of the consensus reaching process, in the NeOn Glossary of Activities two new activities (*Non Ontological Resource Reengineering* and *Non Ontological Resource Reuse*) and their corresponding definitions were included. Furthermore, based on different linguistic revisions some agreed definitions have been slightly modified. This means that the NeOn Glossary of Activities is a living resource in which new activities and definitions can be introduced and updated. Figure 2 presents 4 activities with their corresponding definitions in the glossary, as they are in the wiki used during the consensus reaching process.

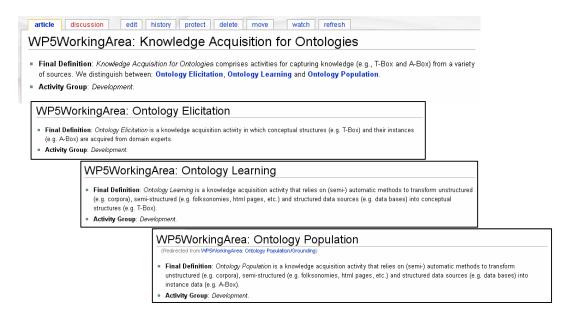


Figure 2 Examples of Activity Definitions

Summarizing, in the NeOn Glossary, some of the activities have maintained their definition taken from the literature. This is the case of, for example: ontology configuration management (Gómez-Pérez et al., 2003), ontology evolution (Stojanovic, 2004), and ontology formalization (Gómez-Pérez et al., 2003). For most of the activities the definition have been changed and adapted based on past definitions existing in the literature and based on NeOn partner comments and discussions. This is the case of, for example, ontology conceptualization, ontology documentation, ontology matching, and ontology selection. And new activities and definitions have been created during the consensus building process, either because the activity did not yet

exist in the literature or because the activity had no definition in the literature. This is the case of the following activities: ontology annotation, ontology comparison, ontology customization, ontology diagnosis, ontology elicitation, ontology enrichment, ontology extension, ontology learning, ontology localization, ontology module extraction, ontology partitioning, non ontological resource reengineering, ontology repair, non ontological resource reuse, ontology search, and ontology upgrade.

The current version of the NeOn Glossary of Activities includes definitions of 53 activities, which can be downloaded from a dedicated web page in the NeOn website⁴. Entries in the NeOn Glossary are arranged alphabetically. Additionally, notes have been added to clarify similar activities, and information about synonymous activities has been also included. As already mentioned, the vocabulary included in the glossary is monolingual (English).

Here we present the NeOn Glossary of Activities, ordered alphabetically.

- Ontology Aligning refers to the activity of finding the correspondences between two or more ontologies and storing/exploiting them. A synonym for this activity is Ontology Mapping.
- Ontology Annotation refers to the activity of enriching the ontology with additional information, e.g. metadata or comments.
- Ontology Assessment refers to the activity of checking an ontology against user requirements, such as usability, usefulness, abstraction, quality, etc.
- Ontology Comparison refers to the activity of finding differences between two or more ontologies or between two or more ontology modules.
- Ontology Conceptualization refers to the activity of organizing and structuring the information (data, knowledge, etc.), obtained during the acquisition process, into meaningful models at the knowledge level according to the ontology specification document. This activity is independent of the way in which the ontology implementation will be carried out.
- Ontology Configuration Management refers to the activity of recording all the versions of the documentation, software and ontology code, and of controling the changes.
- □ *Control* refers to the activity of guaranteeing that scheduled activities in the ontology development process are completed in the manner intended to be performed.

^{4.} http://www.neon-project.org/web-content/images/Publications/neonglossaryofactivities.pdf

- Ontology Customization refers to the activity of adapting an ontology to a specific user's needs.
- Ontology Diagnosis refers to the activity of identifying parts of the ontology directly responsible for incorrectness and incompleteness.
 Ontology diagnosis is triggered by ontology validation.
- Ontology Documentation refers to the collection of documents and explanatory comments generated during the entire ontology building process.

Note: Examples of documents external to the implemented ontology include ontology specification documents, sources used for acquiring knowledge, ontology conceptualization document, design and decision criteria, ontological commitments, etc.

Information inside the implemented ontology includes natural language comments, ontology metadata, and implementation code.

In summary: anything that could be useful to help users, who did not build the ontology, to understand and learn how the ontology was built. Note that the level of granularity of descriptions (might) help or hinder the understanding of the ontology.

- Ontology Elicitation is a knowledge acquisition activity in which conceptual structures (e.g. T-Box) and their instances (e.g. A-Box) are acquired from domain experts.
- Ontology Enrichment refers to the activity of extending an ontology with new conceptual structures (e.g., concepts, roles, axioms, etc.).
- Ontology Environment Study refers to the activity of analyzing the environment in which the ontology is going to be developed.
- Ontology Evaluation refers to the activity of checking the technical quality of an ontology against a frame of reference.
- Ontology Evolution refers to the activity of facilitating the modification of an ontology by preserving its consistency.

Note: Ontology Evolution can be seen as a consequence of different activities during the development of the ontology.

- Ontology Extension is an ontology enrichment activity for stretching the ontology in width.
- Ontology Feasibility Study refers to the activity of answering questions like: is it possible to build the ontology? Is it suitable to build the ontology?, etc.

- Ontology Formalization refers to the transformation of a conceptual model into a formal or semi-computable model according to a knowledge representation paradigm (e.g., description logics, frames, rules, etc.).
- Ontology Forward Engineering refers to the activity of outputting a new implementation of the ontology on the basis of the new conceptual model.
- □ *Ontology Implementation* refers to the activity of generating computable models according to the syntax of a formal representation language (e.g., RDF(S), OWL, F-Logic, etc.).
- Ontology Integration refers to the activity of including one ontology in another ontology.
- □ *Knowledge Acquisition for Ontologies* comprises activities for capturing knowledge (e.g., T-Box and A-Box) from a variety of sources (e.g., documents, experts, data bases, etc.). We can distinguish between: Ontology Elicitation, Ontology Learning and Ontology Population.
- Ontology Learning is a knowledge acquisition activity that relies on (semi-) automatic methods to transform unstructured (e.g. corpora), semi-structured (e.g. folksonomies, html pages, etc.) and structured data sources (e.g. data bases) into conceptual structures (e.g. T-Box).
- Ontology Localization refers to the adaptation of an ontology to a particular language and culture.
- Ontology Mapping refers to the activity of finding the correspondences between two or more ontologies and storing/exploiting them. A synonym for this activity is Ontology Aligning.
- Ontology Matching refers to the activity of finding or discovering relationships or correspondences between entities of different ontologies or ontology modules.
 - *Note*: Ontology Matching can be seen as the first stage of Ontology Aligning.
- Ontology Merging refers to the activity of creating a new ontology or ontology module from two or more, possibly overlapping, source ontologies or ontology modules.
- Ontology Modification refers to the activity of changing the ontology, without considering the consistency.
- Ontology Modularization refers to the activity of identifying one or more modules in an ontology with the purpose of supporting reuse or maintenance.

Note: We can make distinctions between: Ontology Module Extraction and Ontology Partitioning.

- Ontology Module Extraction refers to the activity of obtaining from an ontology concrete modules to be used for a particular purpose (e.g. to contain a particular sub-vocabulary of the original ontology).
- Ontology Partitioning refers to the activity of dividing an ontology into a set of (not necessary disjoint) modules that together form an ontology and that can be treated separately.
- Ontology Population is a knowledge acquisition activity that relies on (semi-) automatic methods to transform unstructured (e.g. corpora), semi-structured (e.g. folksonomies, html pages, etc.) and structured data sources (e.g. data bases) into instance data (e.g. A-Box).
- Ontology Pruning refers to the activity of discarding conceptual structures (e.g., part of T-Box) of a given ontology that are not or no longer relevant.

Note: Pruning is mostly used in combination with ontology learning methods to discard potentially irrelevant learned concepts/relations.

- □ *Ontology Quality Assurance* refers to the activity of assuring that the quality of each and every process carried out and product built (ontology, software and documentation) is satisfactory.
- □ *Non Ontological Resource Reengineering* refers to the process of retrieving and transforming an existing non ontological resource (data bases, controlled vocabularies, etc.) into an ontology.
- Ontology Reengineering refers to the process of retrieving and transforming a conceptual model of an existing and implemented ontology into a new, more correct and more complete conceptual model which is reimplemented.
- Ontology Restructuring refers to the activity of correcting and reorganizing the knowledge contained in an initial conceptual model, and detecting missing knowledge.

Note: This process contains two phases: analysis and synthesis. The "analysis phase goal" is to evaluate the ontology technically, that is, to check that the hierarchy of the ontology and its classes, instances, relations and functions are complete (contain all the definitions required for the domain of chemical substances), consistent (there are no contradictions in the ontology and with respect to the knowledge sources used), concise (there are no explicit and implicit redundancies) and

syntactically correct. The "synthesis phase" seeks to correct the ontology after the analysis phase and document any changes made.

- □ *Ontology Repair* refers to the activity of resolving errors (incompleteness, incorrectness) in the ontology and it is triggered by ontology diagnosis.
- □ *Non Ontological Resource Reuse* refers to taking available non ontological resources (data bases, controlled vocabularies, etc.) for the development of ontologies.
- Ontology Reuse refers to using an ontology or an ontology module in the solution of different problems. Ontology reuse is the activity that allows employing an ontology or an ontology module in, for example, the development of new ontologies, the development of different ontology-based applications, the activity of ontology aligning (as background knowledge), etc.
- Ontology Reverse Engineering refers to the activity of outputting a
 possible conceptual model on the basis of the code in which the ontology
 is implemented.
- □ **Scheduling** refers to the activity of identifying the tasks to be performed during the ontology development, their arrangement, and the time and resources needed for their completion.
- Ontology Search refers to the activity of finding candidate ontologies or ontology modules to be reused.
- Ontology Selection refers to the activity of choosing the most suitable ontologies or ontology modules among those available in an ontology repository or library, for a concrete domain of interest and associated tasks.
- □ *Ontology Specialization* is an ontology enrichment activity for extending the ontology in depth.
- Ontology Specification is a collection of requirements that the ontology should fulfill, e.g. reasons to build the ontology, target group, intended uses, possibly reached through a consensus process.
- Ontology Summarization refers to the activity of providing an abstract or summary of the ontology content.

Note: The summary can include, for example, a couple of top levels in the ontology class hierarchy (perhaps a graphical representation of these top-level concepts and links between them).

 Ontology Translation refers to the activity of changing the representation formalism or language of an ontology from one to another.

Note: Ontology Translation can be part of an ontology reengineering process.

- Ontology Update refers to minor changes carried out in an ontology that could not be considered an upgrade.
- Ontology Upgrade refers to the activity of replacing an existing ontology with a new version.
- Ontology Validation is the ontology evaluation that compares the meaning of the ontology definitions against the intended model of the world aiming to conceptualize.

Note: It answers the question: Are you producing the right ontology?

Ontology Verification is the ontology evaluation which compares the ontology against the ontology specification document (ontology requirements and competency questions), thus ensuring that the ontology is built correctly (in compliance with the ontology specification).

Note: It answers the question: Are you producing the ontology right?

• Ontology Versioning refers to the activity of handling ontology changes by creating and managing different versions of the ontology.

5. Conclusions

The consensus reaching process for building the NeOn Glossary of Activities has been presented in this paper. The NeOn Glossary of Activities, which identifies and defines the activities potentially involved in the ontology network construction, is a first step for solving the lack of a standard glossary in the Ontology Engineering in contrast with the *IEEE Standard Glossary of Software Engineering Terminology* (IEEE, 1990) in the Software Engineering field.

We used the wiki technology for reaching agreement on the name of the activities and on their definitions. But we realized that this technology was not enough for supporting the consensus reaching process because at some point it was no easy to finish the discussion and to fix the labels for certain activities and their definitions. For this reason, we decided to use also meetings and e-mails for clarifying some points in particular discussions.

We already published the NeOn Glossary of Activities in the NeOn website and we still delve into the idea of obtaining feedback from the Ontology Engineering community (outside NeOn). For this purpose, we plan to use the argumentation tool Cicero, built on top of the wiki technology, so that the Ontology Engineering community will can comment the activity definitions for about a year. The long term goal is to have a more complete and consensuated glossary, which could become the terminological reference in the Ontology Engineering field.

Finally, if the Ontology Engineering community supports the activity of providing feedback on the NeOn Glossary of Activities, then we could think of approaching standardization agencies such as IEEE, ISO or W3C for the standardization of the NeOn Glossary of Activities.

Acknowledgments. This work has been supported by the NeOn project (IST-2005-027595). We are very grateful to our NeOn partners and to Rosario Plaza, Guadalupe Aguado de Cea, and Elena Montiel-Ponsoda for their revisions and comments.

References to Publications and Bibliography

IEEE Std 610.12-1990; *IEEE Standard Glossary of Software Engineering Terminology*. 1990.

ISO 12200:1999. Computer applications in terminology. Machine-readable terminology interchange format (MARTIf). Negotiated interchange. 1999.

ISO 1087-1:2000. Terminology work. Vocabulary. Part 1: Theory and application. 2000.

NeOn Consortium; NeOn: Lifecycle Support for Networked Ontologies. Technical Annex. 2006.

Fernández-López, Mariano; Gómez-Pérez, Asunción; Juristo, Natalia; *METHONTOLOGY: From Ontological Art Towards Ontological Engineering*. Spring Symposium on Ontological Engineering of AAAI. Stanford University, California, pp 33–40. 1997.

Gómez-Pérez, Asunción; Fernández-López, Mariano; Corcho, Oscar; *Ontological Engineering*. Springer Verlag. Advanced Information and Knowledge Processing series. ISBN 1-85233-551-3. November 2003.

Kalfoglou, Yannis & Schorlemmer, Marco; *Ontology mapping: the state of the art*. The Knowledge Engineering Review. Volume 18, Issue 1. Pages: 1-31. ISSN:0269-8889. January 2003.

Kotis, Konstantinos & Vouros, George; *The HCONE Approach to Ontology Merging*. In: The Semantic Web: Research and Applications. First European Semantic Web Symposium (ESWS 2004). Pp. 137-151. 2004.

Landau, Sidney I.; *Dictionaries. The art and craft of lexicography*. Cambridge: Cambridge University Press. 1984.

Leuf, Bo & Cunningham, Ward; *The Wiki Way*. Addison-Wesley, 2001.

Stojanovic, Ljiljana; *Methods and Tools for Ontology Evolution*. Dissertation. Referee: Rudi Studer, Christof Weinhardt, Asunción Gómez-Pérez. 2004.

Stojanovic, Ljiljana; Stojanovic, Nenad; Handschuh, Siegfried; *Evolution of the Metadata in the Ontology-based Knowledge Management Systems*. 1st German Workshop on Experience Management: Sharing Experiences about the Sharing of Experience. Berlin. March 7-8, 2002.

Suárez-Figueroa, Mari Carmen & Gómez-Pérez, Asunción; *Towards a Glossary of Activities in the Ontology Engineering Field*. In Proceedings of 6th International Conference on Language Resources and Evaluation (LREC'08), Marrakech. 2008.

Viégas, Fernanda B.; Wattenberg, Martin; Dave, Kushal; *Studying cooperation and conflict between authors with history flow visualizations*. Conference on Human Factors in Computing Systems 2004. Vienna, Austria. Pages. 575-582. ISBN:1-58113-702-8. 2004.