Linköping Studies in Science and Technology

Dissertation No. 1110

Aligning Biomedical Ontologies

by

He Tan

National Graduate School of Computer Science

Department of Computer and Information Science
Linköpings universitet
SE-581 83 Linköping, Sweden

Linköping 2007

Abstract

The amount of biomedical information that is disseminated over the Web increases every day. This rich resource is used to find solutions to challenges across the life sciences. The Semantic Web for life sciences shows promise for effectively and efficiently locating, integrating, querying and inferring related information that is needed in daily biomedical research. One of the key technologies in the Semantic Web is ontologies, which furnish the semantics of the Semantic Web. A large number of biomedical ontologies have been developed. Many of these ontologies contain overlapping information, but it is unlikely that eventually there will be one single set of standard ontologies to which everyone will conform. Therefore, applications often need to deal with multiple overlapping ontologies, but the heterogeneity of ontologies hampers interoperability between different ontologies. Aligning ontologies, i.e. identifying relationships between different ontologies, aims to overcome this problem.

A number of ontology alignment systems have been developed. In these systems various techniques and ideas have been proposed to facilitate identification of alignments between ontologies. However, there still is a range of issues to be addressed when we have alignment problems at hand. The work in this thesis contributes to three different aspects of identification of high quality alignments: 1) *Ontology alignment strategies and systems.* We surveyed the existing ontology alignment systems, and proposed a general ontology alignment framework. Most existing systems can be seen as instantiations of the framework. Also, we developed a system for aligning biomedical ontologies (SAMBO) according to this framework. We implemented various alignment strategies in the system. 2) *Evaluation of ontology alignment strategies.* We developed and implemented the KitAMO framework for comparative evaluation of different alignment strategies, and we evaluated different alignment strategies using the implementation. 3) *Recommending optimal alignment strategies for different applications.* We proposed a method for making recommendations.

Acknowledgments

I would like to express my deepest gratitude to my supervisor Professor Patrick Lambrix, who has introduced me to research. Thank you for your invaluable guidance, discussions and comments during the work. I appreciate your encouragement and support. Your steady, reliable and attentive presence provides such a comfortable work environment. I would like to express my appreciation to Professor Nahid Shahmehri for providing useful comments and pointing out important aspects of the research world.

I am thankful to my colleagues at the Laboratory for Intelligent Information Systems (IISLAB) and at the Division for Database and Information Techniques (ADIT).

Thanks to my friends. I appreciate the times we share challenges and celebrate joys of life. Last but not least, I feel grateful to my parents. Thank you for your understanding, support and endless love. The warmth you give is the thing I cherish most in the world.

We acknowledge the financial support of the Swedish Research Council (Vetenskapsrådet), the Swedish National Graduate School in Computer Science (CUGS), and of the EU Network of Excellence REWERSE (Sixth Framework Programme project 506779).

Tan, He
Linköping, Sweden
June, 2007

List of Enclosed Papers

This thesis contains revised versions of the following papers:

1. P. Lambrix, H. Tan, SAMBO - A System for Aligning and Merging Biomedical Ontologies, *Journal of Web Semantics*, Special issue on semantic web for the life sciences, 2006, 4(3):196-206.
2. H. Tan, V. Jakonienė, P. Lambrix, J. Aberg, N. Shahmehri, Alignment of biomedical ontologies using life science literature, *Proceedings of the International Workshop on Knowledge Discovery in Life Science Literature*, LNBI 3886, Singapore, 2006, 1-17.
3. B. Chen, H. Tan, P. Lambrix, Structure-based filtering for ontology alignment, *Proceedings of the IEEE WETICE Workshop on Semantic Technologies in Collaborative Applications*, 2006, 364-369.
4. P. Lambrix, H. Tan, A Tool for Evaluating Ontology Alignment Strategies, *Journal on Data Semantics*, LNCS 4380, 2007, VIII:182-202.
5. H. Tan, P. Lambrix, A Method for Recommending Ontology Alignment Strategies.

List of Additional Papers

Related Papers

The following are related research articles not included in the thesis.

1. P. Lambrix, A. Edberg, C. Manis, H. Tan, Merging DAML+OIL bio-ontologies, *Proceedings of the International Workshop on Description Logics*, Rome, Italy, 2003.
2. P. Lambrix, H. Tan, Merging DAML+OIL Ontologies, *J. Barzdins, A. Caplinskas, (eds.)*, *Databases and Information Systems (Selected Papers from the Sixth International Baltic Conference DB&IS'2004)*, IOS Press, 2005, 249-258.
3. P. Lambrix, H. Tan, A Framework for Aligning Ontologies, *Proceedings of the Third Workshop on Principles and Practice of Semantic Web Reasoning*, LNCS 3703, Dagstuhl, Germany, 2005, 17-31.
4. P. Lambrix, H. Tan, Ontology alignment and merging, *C. Burger, D. Davidson, R. Baldock, (eds.)*, *Anatomy Ontologies for Bioinformatics: Principles and Practice*, Springer, 2007.
5. T. Wächter, H. Tan, A. Wobst, P. Lambrix, M. Schroeder, A Corpus-driven Approach for Design, Evolution and Alignment of Ontologies, *Proceedings of the Winter Simulation Conference*, 2006, 1595-1602. Invited contribution.

Other published articles

1. P. Lambrix, H. Tan, V. Jakoniene, L. Strömbäck, Biological Ontologies, *C. Baker, K.H. Cheung, (eds.)*, *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*, Springer, 2007, 85-99.
2. L. Strömbäck, V. Jakoniene, H. Tan, P. Lambrix, Representing, storing and accessing molecular interaction data: a review of models and tools, *Briefings in Bioinformatics*, 2006, 7(4):331-338.

EU Network of Excellence REWERSE deliverables

1. R. Backofen, A. Burger, A. Busch, G. Dawelbait, F. Fages, V. Jakoniene, P. Lambrix, K. McLeod, S. Soliman, H. Tan, S. Will, Implementation of prototypes, REWERSE Deliverable A2-D6, 2007.

2. R. Backofen, M. Badea, P. Barahona, M. Berndtsson, A. Burger, G. Dawelbait, A. Doms, F. Fages, A. Hotaran, V. Jakoniene, L. Krippahl, P. Lambrix, K. McLeod, W. Nutt, B. Olsson, M. Schroeder, A. Schroiff, S. Soliman, H. Tan, D. Tilivea, S. Will, Requirements and specification of use cases, REWERSE Deliverable A2-D3, 2005.
3. R. Backofen, M. Badea M, P. Barahona, A. Burger, G. Dawelbait, A. Doms, F. Fages, A. Hotaran, V. Jakoniene, L. Krippahl, P. Lambrix, K. McLeod, S. Möller, W. Nutt, B. Olsson, M. Schroeder, S. Soliman, H. Tan, D. Tilivea, S. Will, Usage of bioinformatics tools and identification of information sources, REWERSE Deliverable A2-D2, 2005.

Table of Contents

Introduction

1 Motivation

In recent decades, information in the biomedical area has been growing explosively: the technological revolution has brought vast quantities of data on genomes, enzymes, pathways and diseases. For example, the sequences of genomes for humans have been completed [CMP03, FJTO03]. In the post-genomic era, this information is used to find solutions to challenges across the life sciences, for instance, detecting systemic functional behaviors of the cell and the organism [CGGG03]. Biomedical information is disseminated over a huge number of heterogeneous and autonomous sources that are often available on the Internet. For instance, 968 datasets freely available on the Web are listed in the 2007 database issue of Nucleic Acids Research [Gal07]. Effectively and efficiently locating, integrating, querying and inferring related data and knowledge are difficulties that researchers experience daily in biomedical research [Lam05]. The Semantic Web for life sciences [HCLS, REW] shows great promise to alleviate these difficulties (e.g. [SBL06, BC07]).

The Semantic Web is an extension of the current Web where information is given well-defined meaning, better enabling machines to automatically process information [BHL01]. One of the key technologies in the Semantic Web is ontologies, which will furnish the semantics for the Semantic Web [SBH06]. Ontologies (e.g. [Gom99]) can be seen as defining the basic concepts and relations of a domain of interest, as well as the rules for combining these concepts and relations. They lead to a better understanding of a field and to more effective and efficient handling of information in that field. The benefits of the use of ontologies include reuse and sharing of knowledge across platforms, and improved documentation, maintenance, reliability and interoperability (e.g. [JU99]). A recent example is that the Royal Society of Chemistry (RSC) Publishing has decided to annotate their published scientific literature with the Gene Ontology (GO) [GO] and the Sequence Ontology (SO) [SO] terms, enabling more quick and effective search and automated discovery of relevant research [GORSC].

A large number of biomedical ontologies have been developed, and much effort has been made to advance their use to facilitate handling vast quantities of information in biomedical research [BS06]. For example, the Gene Ontology consortium [GO], Open Biomedical Ontologies (OBO) [OBO], and SNOMED international [SNOMED], are international research cooperations for development and use of biomedical ontologies. The National Center for Biomedical Ontology [NCBO] is a consortium founded at the US National

Institutes of Health (NIH), whose goal is to support biomedical researchers using ontologies in their work.

Ontologies are developed in different communities based on their own needs. It is inevitable that many biomedical ontologies contain overlapping information, but usually there is no straightforward way to map the overlapping regions. For instance, in the domain of anatomy, a number of ontologies have been developed, e.g. Foundational Model of Anatomy (FMA) [FMA], the Anatomical Dictionary for the Adult Mouse (MA) [HMC+05], anatomic categories of the Medical Subject Headings (MeSH) and the U.S. National Cancer Institute (NCI) thesaurus [NCI]. All these anatomical ontologies have a great deal in common, but are represented in different languages, and impose different syntax, structures, semantics, scopes and perspectives. Furthermore, it is unlikely that eventually there will be one single set of standard ontologies to which everyone will conform [NRM04], although much effort has been devoted to developing standard ontologies in the biomedical area, e.g. the Common Anatomy Reference Ontology (CARO) [HN07] being developed under NCBO. Therefore, applications must often deal with multiple ontologies. Applications often need ontologies from different areas, or which express different views on one area, for example, for querying data over a range of sources annotated with terms from different ontologies. Application developers often need to develop custom, task-specific and smaller ontologies and link them to the standard ontologies. Ontology builders may use existing ontologies as the basis for the creation of new ontologies by extending the existing ontologies or by combining knowledge from different smaller ontologies. In all of these cases, interoperability of multiple ontologies is hindered by their heterogeneity, which has become a critical issue for realizing the vision of the Semantic Web [Noy05, BLS07].

Aligning ontologies, i.e. identifying relationships between different ontologies, aims to overcome the problem. Currently, thousands of ontologies have been developed, and billions of sources on the Web are annotated with ontology terms [SWG]. High quality ontology alignment is a cornerstone for seamless interoperability between sources using different ontologies. It is a necessary precondition for effectively locating, integrating, querying and inferring information spread over heterogeneous sources. Our work contributes to solutions for efficiently identifying high quality alignments. In this thesis we focus on three issues: developing tools for aligning ontologies, evaluating ontology alignment strategies, and recommending optimal alignment strategies for different applications. Although our current work mainly contributes to the field of the life sciences which is seen to be a huge area for potential application development of the Semantic Web, the results can be generalized and are applicable to different areas.

2 Background

2.1 Ontologies

Ontology as the study of existence dates back to ancient Greek philosophy. Ontology studies what kinds of things exist, how these things can be interrelated and what they mean in reality. The term has been used in very different ways within information science [GG95]. In this thesis, we follow an early definition provided in the field of the Artificial Intelligence [NFF91]: An ontology defines the basic concepts and relations of a domain of interest, as well as the rules for combining these concepts and relations. The definition gives a rough image of how an ontology is built. According to the definition, an ontology includes not only the terms that are explicitly defined, but also terms that can be inferred using rules.

The benefits of the use of ontologies include reuse and sharing of knowledge across platforms, and improved documentation, maintenance, reliability and interoperability (e.g. [JU99]). Ontologies are used for communication between people and organizations by providing a common terminology over a domain. They provide the basis for interoperability between systems. They can be used for making the content in information sources explicit and serve as an index to a repository of information. Additionally, they can be used as a basis for integration of information sources and as a query model for information sources. They also support clearly separating domain knowledge from application-based knowledge as well as validation of data sources. Overall, ontologies lead to a better understanding of a field and to more effective and efficient handling of information in that field.

Ontologies can be classified into different types depending on which of the components are represented and the kind of information they can represent (e.g [SGB01, LTJS07]). *Controlled vocabularies* are a simple type of ontology. There are essentially lists of concepts that are enumerated explicitly. When the concepts are organized into a hierarchical structure, we obtain a *taxonomy*. The relations in a taxonomy can be *is-a* and *part-of*. *Thesauri* are more complex ontologies, in which there are also relations that connect concepts across the hierarchy. Most currently available biomedical ontologies are of these three types of ontologies. Other types of ontologies are *data models* and *knowledge bases.* In *data models* axioms – which describe facts that are always true in topic area of the ontology – are used in addition to concepts and relations. *Knowledge bases* are often based on a logic. In this thesis we focus on *controlled vocabularies*, *taxonomies* and *thesauri.*

Ontology languages allow users to represent an ontology. The main requirements of ontology languages are [AH03]: 1, a well-defined syntax, 2, well-defined semantics, which is a necessary condition for precisely describing the meaning of knowledge, 3, efficient reasoning support, which can be used to, e.g. check the consistency of the ontology, detect unsatisfiable of concepts and remove redundancy. 4, sufficient expressive power, which determines what kinds of knowledge can be represented in the ontology, and 5, convenience of expression. There are a number of languages, e.g. RDF(s) [RDF], F-Logic [KLW95] and OWL [OWL]. Few of the existing ontology languages fulfill all of these requirements.

2.2 Heterogeneity of ontologies

The heterogeneity of ontologies may occur at different levels (e.g. [Kle01, BEF04]). At *language level*, ontologies are represented in different languages, such as RDF(s) and OWL. At *terminology level*, different namings of terms bring out the mismatching between ontologies, for example, synonyms, homonyms and different encodings. At *concept level*, we encounter the heterogeneity of ontologies due to different scope, granularity and perspective. At *pragmatics level*, knowledge represented in ontologies may be in different context.

The focus of our work is to overcome the heterogeneity of ontologies on the terminology and concept levels. *Aligning*, *mapping* and *articulating ontologies* are the terms often used to refer to the task, and different meanings of the terms are given in the literature. In this thesis we use **aligning ontologies**. It is defined as identifying the relationships between concepts or relations from two different source ontologies. The relationships can be equivalence as well as *is-a*, *part-of* or any other kind of relation. This definition gives a picture of how to specify correspondence between two ontologies. This definition covers the common meaning of *mapping ontologies. Aligning ontologies* usually is to bring together a set of ontologies, and *articulating ontologies* is sometimes defined as aligning parts of ontologies. We also include **merging ontologies** in this thesis, which is to derive a new ontology which is a (minimal) union of two existing ontologies based on their identified alignments.

Translating ontology is the common term referring to tackling heterogeneity on the language level, which is defined as converting an ontology represented in one ontology language to another language while preserving the semantics. Some tools have been developed for facilitating this process, e.g. Jena [Jena] is a widely used tool which supports ontologies in RDFs, DAML+OIL [DAML] and OWL. In the biomedical area OWL has been widely accepted, e.g. in OBO and NCBO. We do not focus on this issue, and assume that ontologies to be aligned are in OWL. The pragmatics heterogeneity is related to the use of ontologies. There is not much work on this issue.

2.3 Aligning Ontologies

A number of systems for aligning ontologies have been developed, and various methods for discovering alignments are proposed in these systems. Most of the existing systems are semi-automatic, i.e. a user is involved in identifying alignments. To detect alignments, in such a system there are usually matchers which compute similarity values between terms from different ontologies, and algorithms which filter out alignment suggestions based on results from matchers. If there are several matchers, algorithms may be available to combine their results. So, alignment strategies are composed of different **matchers**, **combination** and **filtering algorithms**. Further, in such a system there are interactive components where the user decides alignments based on suggestions.

3 Problem Statement

Aligning ontologies becomes a critical issue when collaborating applications use different ontologies. A lot of research is currently dealing with this challenge. A number of ontology alignment systems have been developed. In these systems various techniques and ideas have been proposed to facilitate identifying alignments between ontologies. However, there is still a range of issues to be addressed when we have alignment problems at hand.

No alignment system is suitable for all kinds of alignment problems. Developers may need to build an alignment system from scratch for their particular need. For example, there is no (semi-)automatic alignment tool in particular for ontologies from the field of life sciences. However, it is not clear how such systems should be developed and what would be a suitable architecture for such a system. Further, in the field of life sciences domain knowledge is available in the form of biomedical domain lexicons, databases and collections of articles. This knowledge could be utilized for finding high quality alignments, but not much work has been done on this. So, the first problem to be tackled is in this thesis is:

> **Problem 1:** There are very few guidelines on building alignment systems.

Although most of alignment systems and strategies are developed for their own need, many of them can be extended for other applications. The evaluation and comparison of different systems and their alignment strategies would give us their strengths and weaknesses, and valuable insight into their properties. It would lead to improvement of the existing alignment systems and strategies, but also provide valuable knowledge when they are reused for other alignment problems. Thus the second problem is that,

> **Problem 2:** Up to now there are few comparative evaluations, and also there are very few tools to support evaluations.

In most applications where alignments between multiple ontologies are required, domain experts want to reuse the existing alignment strategies. It is unlikely that there is one single strategy that outperforms the others for all different kinds of ontologies, therefore domain experts must choose appropriate strategies from different available strategies. It is a very difficult task.

> **Problem 3**: Currently, very little knowledge is available about existing alignment methods, their applicability, and the ontologies to be aligned, and yet no tool exists that facilitates the task, e.g. by recommending appropriate alignment strategies.

As discussed in section 2, our current work will focus on ontologies which are *controlled vocabularies*, *taxonomies* and *thesauri*. We assume that ontologies to be aligned are in OWL, and we do not consider the context of ontologies in application. Further, we focus on semi-automatic alignment systems.

4 Contributions

This thesis contributes to aligning biomedical ontologies as follows.

4.1 Ontology alignment systems and strategies

- As a first step in dealing with problem 1, we adapted the method for the classification of automatic schema matching approaches in [RB01] and used it for classification of ontology alignment approaches. We classified ontology alignment approaches based on the knowledge they use. The classification augments our understanding of alignment strategies. An overview of the existing ontology alignment systems and the strategies they use is given in paper 1.

- As the second step in coping with problem 1, we proposed, based on contribution 1, a general framework for aligning ontologies. The framework provides a basis for building alignment systems, and provides support for experimenting with different alignment components and their combinations. Part of the framework is similar to some steps in the alignment process introduced in [ES04b], but our framework focuses on an architecture for alignment systems. Most of the existing systems can be seen as an instantiations of our framework. This contribution is reported in paper 1.

- We developed SAMBO (System for Aligning and Merging Biomedical Ontologies) according to the general framework. Currently, it is the only alignment tool targeted towards biomedical ontologies. We studied the characteristics of the existing biomedical ontologies and sought available sources in the area, which can be utilized for alignment. We implemented several alignment algorithms, filtering algorithms and a weighted sum combination algorithm. We developed and implemented strategies based on the textual descriptions of the terms, *is-a* and *part-of* hierarchies of the ontologies, retrieved instances and use of thesauri. In paper 2 we report on algorithms that utilize life science literature to compute the similarity between terms. In paper 3 we propose an approach that applies information about the structure of ontologies in the filtering phase. All other implemented alignment strategies are discussed in paper 1.

4.2 Evaluation of alignment systems and strategies

- As the first step in dealing with problem 2, we evaluated the different alignment strategies as well as their combinations within SAMBO. We also compared SAMBO with other ontology alignment systems. In the evaluation we focused on the quality of the suggestions which is measured using precision and recall. We use precision as it is usually defined in information retrieval, i.e. the number of correct suggestions divided by the number of suggestions. Similarly, recall is defined as the number of correct suggestions divided by the total number of correct suggestions, in this case the expected suggestions. Our work is one of the two existing evaluations which focus on the quality of the suggestions. In addition, we studied the evaluation results to detect strengths and weaknesses of different strategies and their combinations. The results of the evaluations are discussed in paper 1.

- Currently, there is no tool that provides an integrated environment for comparative evaluations on alignment strategies and their combinations. The paper 4 contributes to this issue. Based on our general alignment framework and experiments with SAMBO, we proposed the KitAMO framework which is an integrated system to comparatively evaluate non-interactive alignment components and their combinations. We implemented a prototype according to the framework. With the implementation we experimented with comparative evaluations of different alignment strategies and their combinations in terms of the performance and the quality of the alignment suggestions. We also studied how the evaluation results can be used to detect strengths and weaknesses of different strategies and their combinations.

4.3 Recommendation on alignment strategies

- Up to now two methods have been proposed to deal with problem 3. They require prior knowledge about the alignment problem and existing alignment strategies. However, such knowledge is usually not available when making recommendations. In our work we proposed a method in which the prior knowledge is not necessary, and which also minimizes the effort from users. We discussed the method in the setting of an alignment problem with two biomedical ontologies in paper 5.

5 Research Methods

Survey In the process of understanding the field of ontology alignment, we surveyed the scientific publications on ontology alignment and a related area, schema matching. We experimented with publicly available alignment tools using a set of biomedical ontologies. We studied their functionalities and quality of alignment. The survey results were used as a basis for the classification of alignment techniques and systems, and for a general alignment framework.

Prototyping Prototyping leads to a better and deeper understanding of research problems and their proposed solutions. We developed our alignment system based on the general framework, and implemented different alignment strategies in the system. We performed a number of evaluations and comparisons on alignment strategies and systems. The experiments demonstrated the applicability of the general framework, and gave a better understanding of the alignment framework. In addition, it formed a basis for the work on problem 2 and 3.

When dealing with problem 2, we developed a prototype for evaluating alignment strategies. We experimented with our implemented strategies to demonstrate and evaluate the applicability of the prototype. The experiments also indicated that the prototype could be applied to support the recommendation method we proposed for problem 3. We implemented several algorithms in the recommendation method. We tested the applicability of the method and evaluated the quality of the recommendations when different algorithms were used. This experiment could, in its turn, be seen as an application study of the evaluation prototype.

6 Paper Summaries

In this section we give short summaries of the five papers enclosed in this thesis. In paper 1, 2, and 3 we investigate some fundamental problems of aligning ontologies. Paper 4 deals with the comparative evaluation of the different alignment strategies. In paper 5 we approach making recommendations on alignment strategies for a particular alignment problem.

Paper 1: SAMBO - A System for Aligning and Merging Biomedical Ontologies

In this paper we present a general framework for aligning ontologies. In the framework the alignments between two source ontologies can be determined under the supervision of the user based on the suggestions generated by different alignment strategies as well as their combinations. Further, we identify the types of alignment strategies used in the existing alignment systems. We describe SAMBO, an alignment system which exemplifies the use of the framework. In SAMBO we implement several matchers that are representatives of the different types of alignment strategies, a filtering and a combination algorithm. We discuss the evaluations in which we compare the quality of suggestions generated by the different matchers and their combinations with the filtering algorithm, and the amount of time they take to generate the suggestions. We also compare SAMBO with two other alignment systems with respect to the quality of suggestions. We use several well-known biomedical ontologies in the evaluations.

Paper 2: Alignment of biomedical ontologies using life science literature

In this paper we present our experiments on utilizing life science literature related to concepts from the source ontologies. We propose a basic algorithm as well as extensions that take the structure of the ontologies into account. The algorithms build on the intuition that a similarity measure between terms can be computed based on the probability that documents about one term are also about the other term and vice versa. The related literature is retrieved from PubMed [PubMed], a free digital archive of biomedical and life sciences journal literature. We compare the basic matcher with its extensions, and with other alignment strategies implemented in SAMBO in terms of the quality of suggestions and the amount of time they take to generate the suggestions. We also present the results of evaluations on the different combinations of these matchers with other implemented alignment strategies.

Paper 3: Structure-based filtering for ontology alignment

In this paper we propose a method where we use the structural information in the filtering stage in the alignment process. The approach is based on the observation that the precision of the results usually decreases but the recall increases when the thresholds decreases in the threshold filtering algorithm. To keep the precision and include more correct suggestions, the approach augments the threshold filtering algorithm with the structural information. We apply the algorithm to filter out the suggestions from several matchers implemented in SAMBO. We compare the quality of the suggestions filtered out with and without the structure information, and also compare the amount of time for computation of the suggestions and filtering.

Paper 4: A Tool for Evaluating Ontology Alignment Strategies

In paper 4 we propose the KitAMO framework for comparative evaluation of different ontology alignment strategies and their combinations. We present our current implementation of the framework. With the implementation, we illustrate how the system can be used to evaluate and compare alignment strategies and their combinations in terms of performance and quality of the proposed alignments, and give suggestions on how the results can be analyzed to obtain deeper insights into the properties of the strategies.

Paper 5: A Method for Recommending Ontology Alignment Strategies

In this paper we propose a method that provides recommendations on alignment strategies for a given alignment problem. The method makes the recommendations based on the evaluation of the different available alignment strategies on several small selected pieces from the ontologies. In the paper we describe the basic steps of the method, and then illustrate and discuss the method in the setting of an alignment problem with two well-known biomedical ontologies.

7 Related Work

7.1 Ontology alignment systems and strategies

There are a number of systems devoted to aligning ontologies, such as Chimaera [MFRW00], FCA-Merge [SM01], PROMPT [NM03], IF-Map [YM03], GLUE [DMDH03], QOM [ES04a], COMA++ [ADMR05], SST[ZK06]. Different alignment strategies are implemented in these systems. In paper 1 we classify the various alignment approaches based on the knowledge they use into strategies based on linguistic matching, structure-based strategies, instance-based strategies, constraint-based strategies, strategies using auxiliary information, and their combinations. We give a brief overview of 13 systems and discuss their alignment strategies according to this classification. In [EBB04] alignment approaches are classified into local methods and global methods. Local methods are the matchers based on internal knowledge, which are categorized into, terminological methods, structural methods, extensional (based on instances) methods, and semantic methods (based on models). Global methods include matchers composed of several local methods, combination algorithms, methods involving the ontologies as a whole, learning methods, and user input. 21 systems are discussed according to the classifications.

General alignment frameworks are proposed in works by two other groups. [BEF04] defines the alignment process as a function and discusses various constraints that can be applied on input, parameters and output of the function. However, how to compute the function is not given. [ES04a] introduces an alignment process with six main steps: feature engineering, selection of next search steps, similarity computation, similarity aggregation, interpretation and iteration. Feature engineering is to transform the representation of ontologies into a format other steps can use. Selection of next search steps restricts a search space of candidate alignments to reduce the run-time complexity by reducing the number of candidates alignments. Most existing alignment systems usually check all candidate alignments. Further, [ES04b] gives 17 rules for the computation of similarity. In paper 1 we propose our general alignment framework. Part of the framework is similar to the steps, similarity computation, similarity aggregation, interpretation and iteration in the process described in [ES04a]. The main goal of our framework is to give a basis for building alignment systems, and provides support for experimenting with different alignment components and their combinations.

A repository of information on ontology alignment systems can be found on the ontology matching website (http://www.ontologymatching.org/). It is updated frequently.

7.2 Evaluation of alignment systems and strategies

Currently, comparative evaluations of ontology alignment systems have been performed by a few groups. The EU OntoWeb project [OntoWeb] evaluated the systems PROMPT based on Protégé (with extension Anchor-PROMPT), Chimaera (described, not evaluated), FCA-Merge and ODEMerge. This evaluation focused on such things as functionality, interoperability and visualization, but did not include tests on the quality of the alignment suggestions. Our work includes PROMPT, Chimaera, FOAM and SAMBO in the evaluations. The evaluations focus on the quality of the alignment suggestions as well as the time it takes to generate suggestions with these tools. In addition, the interface of the systems are discussed in [LE03], and in paper 1 the different alignment algorithms and their combinations with different combination weights were evaluated with different filtering threshold values. Since 2004, the Ontology Alignment Evaluation Initiative (OAEI, http://oaei.ontologymatching.org/) has organized yearly international evaluation campaigns. The 2006 campaign has 4 tracks gathering 6 data sets: a comparison track, an expressive ontologies track, a directories and thesauri track, and a consensus workshop. Each track has a different evaluation purpose. 10 systems participated in the campaign and 3 of the systems used all the data sets. The campaign focused on the quality of the alignment suggestions. The results are reported at the international workshop on ontology matching (OM-2006).

Further, tools for evaluations are developed in the latter two projects. The Alignment API [Euz06] is the tool developed in the OAEI, which the participants may use in the campaign. The Alignment API includes several evaluators which can compute the precision, recall, fallout and f-measure of an alignment result and a weighted symmetric difference between two alignments. OLA [ELTV04] is a GUI application implemented on top of this API. We propose the KitAMO framework (paper 4) which provides an integrated system for comparative evaluation and analysis of alignment strategies and their combinations.

7.3 Recommendation on alignment strategies

Currently, very little research tackles the problem of selecting alignment strategies that are optimal for given alignment problems. In [MJE06] it is argued that finding appropriate alignment strategies should be based on knowledge about the strategies and their previous use. As a first step a number of factors (related to input, output, approach, usage, cost and documentation) were identified that are relevant when selecting an alignment strategy. The relevant data is collected by questionnaires. This method requires much effort from many different users. The Analytic Hierarchy Process is used to detect suitable alignment approaches. In [ESS05], APFEL, a machine learning approach to optimize alignment strategies, is proposed. In

APFEL a set of feature parameters are declared for the source ontologies, the similarity assessment, and the different matchers, combination and filter algorithms. To generate training data, an existing parametrization is used and alignment suggestions are generated. These suggestions need to be validated by the user. A machine learning approach is then used to learn an optimal parametrization. The main difficulty in this approach is choosing the training data. The validation of data also requires effort from users that have good knowledge about the area of the ontologies. In paper 5 we have proposed and tested a method that is based on the assumption that strategies that work well for aligning selected small segments of the ontologies will also work well for aligning the whole ontologies. The recommendation is made based on two well-known measurements of quality of alignment suggestions.

8 Conclusion

8.1 Summary

Aligning ontologies aims to overcome the heterogeneity of ontologies. High quality ontology alignment is critical for seamless interoperability between sources using different ontologies. The goal of our work is to find solutions for efficiently identifying high quality alignments between ontologies. This thesis contributes to three problems in the field: 1, development of ontology alignment strategies and systems, 2, evaluation of alignment strategies, and 3, recommending suitable alignment strategies for given alignment problems.

Much research has joined the field of aligning ontologies. On the publication list of the ontology matching website, there are 14 conference and journal papers in 2003, and in 2006 there are 48 papers. A number of various alignment systems and methods have been developed. In this thesis we classified ontology alignment strategies based on the knowledge they use. The classification leads to an easy way to understand the field. We proposed a general framework for aligning ontologies. The framework can provide a basis for building alignment systems, and support for experimenting with different alignment strategies. Most of the existing systems can be seen as an instantiations of our framework. Based on the framework we developed SAMBO, a system for aligning biomedical ontologies. In the system we developed different alignment strategies targeted at biomedical ontologies.

To improve the work in the field, comparative evaluation of different systems and methods has become a significant issue. In this thesis we evaluated different strategies implemented in SAMBO, and compared SAMBO with other systems. The evaluation experiments led to KitAMO, a prototype for comparative evaluations of alignment strategies. We implemented the prototype, and evaluated its applicability. However, there are still many open problems in this issue, such as evaluation criteria and construction of test cases.

The outcome of comparative evaluation can provide valuable input to another challenge, recommending suitable alignment strategies for given alignment problems. Currently, there is very little research on the problem. In this thesis we proposed a method for recommendation of alignment strategies. We implemented different algorithms for the different steps in the method, demonstrated the feasibility of the method, and evaluated the algorithms.

8.2 Future Work

In the future we will continue our work for the three problems as follows.

Evaluation of alignment strategies and systems We have proposed the KitAMO framework for comparative evaluations of the non-interactive alignment components, and implemented a prototype of KitAMO. The current implementation supports evaluation of alignment algorithms. In the future we will extend the system to support evaluation of combination and filtering methods. Currently, evaluation in KitAMO can be based on the quality of alignment suggestions and the time that alignment algorithms need to generate suggestions. The quality of suggestions is measured using the recall and precision. Since the user may need other evaluation criteria and measurement methods for different evaluation purposes, KitAMO will include more commonly used evaluation approaches, e.g. fallout and f-measure, and may allow users to plug-in their own evaluation strategies based on a given API. The current test cases are small pieces from large biomedical ontologies, and they are *taxonomies*. We would like to build a testbed which includes test cases with ontologies of different kinds and sizes, supporting different evaluation purposes. Further, the visualization components of the system can be improved.

Recommendation of alignment strategies We have proposed a method for recommending alignment strategies for a particular alignment problem.

Even though we have shown the feasibility of our method and have obtained good results for the alignment problem in the experiment, it is necessary to perform more experiments with different kinds of ontologies. We also want to investigate the influence of the different choices in the different steps of the method. This includes investigating different segment pair selection algorithms, recommendation measures and recommendation algorithms. It will also be interesting to look at how the ideas from [MJE06] and [ESS05] can be used to augment our approach. For instance, when more knowledge is obtained regarding the different strategies and their previous use (as in [MJE06]), this knowledge could be used as a first step to filter the available strategies and it can be used by the recommendation strategy. Also the optimization approach in [ESS05] may be useful for finding better combinations as well as within the recommendation step. The studies will lead us to define a theoretical framework for recommendation. Finally, we intend to develop a tool that supports the framework by extending the KitAMO system.

SAMBO and ontology alignment strategies In SAMBO we have experimented with several alignment strategies. In the future we will develop and implement new alignment algorithms, for example constraint-based algorithms, and we also want to experiment with different combination and filtering approaches. Using KitAMO, we will evaluate the different alignment strategies. The lessons learned in evaluation would lead to improved strategies. On the other hand, different alignment strategies may produce new requirements on KitAMO. A lot of work has been devoted to the non-interactive alignment components, but aligning ontologies is hardly a fully automatic task. The interactive components need more study. We have worked on integrating an interactive ontology visualization tool into SAMBO, which not only improves the user interface to visualize information, but may also lead to the development of new alignment algorithms. Further, SAMBO will be extended to better support ontologies from outside of the biomedical area. We also envision that after evaluation and recommendation of alignment strategies, KitAMO can be asked to generate an alignment system targeted at a particular alignment problem.

References

[ADMR05] D. Aumueller, H. H. Do, S. Massmann, E. Rahm, Schema and ontology matching with COMA++, *Proceedings of SIGMOD (Demonstration)*, 2005.

[AH03] G. Antoniou, F. Harmelen, Web Ontology Language: OWL, *S. Staab and R. Studer (eds.) Handbook on Ontologies in Information Systems*, Springer-Verlag, 2003, 76-92.

[BC07] Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences, Baker, Christopher J.O.; Cheung, Kei-Hoi (eds.) Springer, 2007.

[BEF04] P. Bouquet, J. Euzenat, E. Franconi, L. Serafini, G. Stamou, S. Tessaris, D2.2.1: Specification of a common framework for characterizing alignment, Technical report, *NoE Knowledge Web project deliverable*, 2004.

[BHL01] T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web, *Scientific American*, 2001, 284:34-43.

[BLS07] J. Blake, P. Lambrix, N. Shah, Ontology Alignment, *Dagstuhl Seminar 07132, Towards Interoperability of Biomedical Ontologies*, Dagstuhl, Germany, 2007.

[BS06] O. Bodenreider, R. Stevens, Bio-ontologies: current trends and future directions, *Briefings in Bioinformatics*, 2006, 7(3): 256-274.

[CGGG03] F.S. Collins, E. Green, A. Guttmacher, M. Guyer, A vision for the future of genomics research, *Nature*, 2003, 422:835-847.

[CMP03] F.S. Collins, M. Morgan, A. Patrinos, The Human Genome Project: Lessons from Large-Scale Biology, *Science*, 2003, 300(5617):286-290.

[DAML] DAML+OIL. http://www.w3.org/Submission/2001/12/.

[DMDH03] A. Doan, J. Madhavan, P. Domingos, A. Halevy, Ontology matching: A machine learning approach, *S. Staab and R. Studer (eds.), Handbook on Ontologies in Information Systems*, Springer, 2003, 397-416.

[EBB04] J. Euzenat, J. Barrasa, P. Bouquet, R. Dieng, M. Ehrig, M. Hauswirth, M. Jarrar, R. Lara, D. Maynard, A. Napoli, G. Stamou,

H. Stuckenschmidt, P. Shvaiko, S. Tessaris, S. van Acker, I. Zaihrayeu, T. L. Bach, D2.2.3: State of the art on ontology alignment, Technical report, *NoE Knowledge Web project deliverable*, 2004.

[ELTV04] J. Euzenat, D. Loup, M. Touzani, P. Valtchev, Ontology alignment with OLA, *Proceedings of the 3rd international Workshop on the Evaluation of Ontology-based Tools*, 2004.

[ES04a] M. Ehrig, S. Staab, QOM - quick ontology mapping, *Proceedings of 3rd International Semantic Web Conference*, LNCS 3298, 2004, 683-697.

[ES04b] M. Ehrig, Y. Sure, Ontology Mapping: An Integrated Approach, *Proceedings of 1st European Semantic Web Symposium*, LNCS 3053, 2004, 76-91.

[ESS05] M. Ehrig, S. Staab, Y. Sure, Bootstrapping Ontology Alignment Methods with APFEL, *Proceedings of the International Semantic Web Conference*, 2005, 186-200.

[Euz06] J. Euzenat, An API for ontology alignment, version 2.1, *Ontology Alignment Evaluation Initiative*, 2006.

[FaCT] FaCT. http://www.cs.man.ac.uk/∼horrocks/FaCT/.

[FJTO03] M.E. Frazier, G.M. Johnson, D.G. Thomassen, C.E. Oliver, A. Patrinos, Realizing the Potential of the Genome Revolution: The Genomes to Life Program, *Science*, 2003, 300(5617):290-293.

[FMA] Foundational Model of Anatomy. http://sig.biostr.washington.edu/projects/fm/.

[Gal07] M. Y. Galperin, The Molecular Biology Database Collection: 2007 update, *Nucleic Acids Research*, 2007, 35(Database issue):D3-D4.

[GG95] N. Guarino, P. Giaretta, Ontologies and Knowledge Bases: Towards a Terminological Clarication, *N. Mars (ed.) Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*, IOS Press, 1995, 25-32.

[GO] The Gene Ontology Consortium, Gene Ontology: tool for the unification of biology, *Nature Genetics*, 2000, 25(1):25-29. http://www.geneontology.org/.

[Gom99] A. Gómez-Pérez, Ontological Engineering: A state of the Art, *Expert Update*, 1999, 2(3):33-43.

[GORSC] http://www.rsc.org/Publishing/Journals/cb/Volume/2007/1/GO-RSC.asp

[HCLS] W3C Semantic Web Health Care and Life Sciences Interest Group. http://www.w3.org/2001/sw/hcls/

[HMC+05] T. Hayamizu, M. Mangan, J. Corradi, J. Kadin and M. Ringwald, The Adult Mouse Anatomical Dictionary: a tool for annotating and integrating data, *Genome Biology*, 2005, 6(3):R29.

[HN07] M.A. Haendel, F. Neuhaus, D.S. Osumi-Sutherland, P.M. Mabee, J.L.V. Mejino, C.J. Mungall, and B. Smith, CARO - The Common Anatomy Reference Ontology, chapter in *A. Burger, D. Davidson, R. Baldock, (eds), Anatomy Ontologies for Bioinformatics: Principles and Practice*, Springer, 2007. To appear.

[Jena] Jena - A Semantic Web Framework for Java, http://jena.sourceforge.net/

[JU99] R. Jasper, M. Uschold, A Framework for Understanding and Classifying Ontology Applications, *Proceedings of the 12th Workshop on Knowledge Acquisition, Modeling and Management*, Banff, Canada, 1999.

[Kle01] M. Klein, Combining and Relating Ontologies: An Analysis of Problems and Solutions, *IJCAI Workshop on Ontologies and Information Sharing*, 2001.

[KLW95] M. Kifer, G. Lausen, J. Wu, Logical Foundations of Object-Oriented and Frame-Based Languages, *Journal of the ACM*, 1995, (42):741-843.

[Lam05] P. Lambrix, Towards a Semantic Web for Bioinformatics using Ontology-based Annotation, *Proceedings of the 14th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises*, 2005, 3-7. Invited talk.

[LE03] P. Lambrix, A. Edberg, Evaluation of ontology merging tools in bioinformatics, *Proceedings of the Pacific Symposium on Biocomputing*, 2003, 8:589-600.

[LT05] P. Lambrix, H. Tan, Merging DAML+OIL Ontologies, *J. Barzdins and A. Caplinskas, (eds.), Databases and Information Systems - Selected Papers from the Sixth International Baltic Conference on Databases and Information Systems*, IOS Press, 2005, 249-258.

[LTJS07] P. Lambrix, H. Tan, V. Jakoniene, L. Strömbäck, Biological Ontologies, *C. Baker, K.H. Cheung, (eds), Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*, 2007, 85-99.

[MeSH] Medical Subject Headings. http://www.nlm.nih.gov/mesh/

[MFRW00] D. McGuinness, R. Fikes, J. Rice, S. Wilder, An Environment for Merging and Testing Large Ontologies, *Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning*, 2000, 483-493.

[MJE06] M. Mochol, A. Jentzsch, J. Euzenat, Applying an Analytic Method for Matching Approach Selection, *Proceedings of the International Workshop on Ontology Matching*, 2006.

[NCBO] The National Center for Biomedical Ontology. http://www.bioontology.org/.

[NCI] The NCI thesaurus. http://ncicb.nci.nih.gov/infrastructure/cacore_-overview/vocabulary.

[NFF91] R. Neches, R. Fikes, T. Finin, T. Gruber, R. Patil, T. Senator, W.R. Swartout, Enabling Technology for Knowledge Sharing, *AI Magazine*, 1991, 12(3):36-56.

[NM03] N.F. Noy, M.A. Musen, The PROMPT Suite: Interactive Tools For Ontology Merging And Mapping, *International Journal of Human-Computer Studies*, 2003, 59(6):983-1024.

[Noy05] N. Noy, Order from Chaos, *ACM Queue Magazine, Special Issue on Semi-Structured Data*, 2005, 3(8):42-49.

[NRM04] N.F. Noy, D.L. Rubin, M.A. Musen. Making Biomedical Ontologies and Ontology Repositories Work, *IEEE Intelligent Systems*, 2004, 19(6):78-81.

[OBO] Open Biomedical Ontologies. http://obo.sourceforge.net/.

[OntoWeb] OntoWeb Consortium, Deliverables 1.3: A survey on ontology tools, and 1.4: A survey on methodologies for developing, maintaining, evaluating and reengineering ontologies, 2002.

[OWL] OWL. http://www.w3.org/2004/OWL/.

[Pellet] Pellet. http://pellet.owldl.com/.

[PubMed] PubMed. http://www.ncbi.nlm.nih.gov/entrez/.

[Racer] Racer. http://www.cs.concordia.ca/∼haarslev/racer/.

[RB01] E. Rahm, P. A. Bernstein, A survey of approaches to automatic schema matching, *The VLDB Journal*, 2001, 10:334-350.

[REW] REWERSE EU Network of Excellence on Reasoning on the Web with Rules and Semantics, Deliverables of the A2 Working Group. http://www.rewerse.net.

[RDF] RDF. http://www.w3.org/RDF/.

[SBH06] N. Shadbolt, T. Berners-Lee, W. Hall, The Semantic Web Revisited, *IEEE Intelligent Systems*, 2006, 21(3):96-101.

[SBL06] R. Stevens, O. Bodenreider, Y. A. Lussier, Session introduction: Semantic Webs for Life Sciences, *Proceedings of Pacific Symposium on Biocomputing*, 2006, 11:112-115.

[SGB01] R. Stevens, C. Goble, S. Bechhofer, Ontology-based knowledge representation for bioinformatics, *Briefings in Bioinformatics*, 2001, 1(4):398-414.

[SM01] G. Stumme, A. Mädche, FCA-Merge: Bottom-up merging of ontologies, *Proceedings of the 17th International Joint Conference on Artificial Intelligence*, 2001, 225-230.

[SNOMED] SNOMED-CT. http://www.nhsia.nhs.uk/snomed/pages.

[SO] The Sequence Ontology. http://www.sequenceontology.org/.

[SOFG] SOFG, Standards and Ontologies for Functional Genomics. http://www.sofg.org/.

[SWG] Semantic Web Search (SWOOGLE). http://swoogle.umbc.edu/.

[YM03] Y. Kalfoglou, M. Schorlemmer, IF-Map: an ontology mapping method based on information flow theory, *Journal on data semantics*, 2003, I:98-127.

[ZK06] P. Ziegler, C. Kiefer, C. Sturm, K. R. Dittrich, A. Bernstein, Detecting Similarities in Ontologies with the SOQA-SimPack Toolkit, *Proceedings of International Conference on Extending Database Technology*, 2006, 59-76.

Department of Computer and Information Science
Linköpings universitet

Dissertations

Linköping Studies in Science and Technology

No 14 **Anders Haraldsson:** A Program Manipulation System Based on Partial Evaluation, 1977, ISBN 91-7372-144-1.

No 17 **Bengt Magnhagen:** Probability Based Verification of Time Margins in Digital Designs, 1977, ISBN 91-7372-157-3.

No 18 **Mats Cedwall:** Semantisk analys av processbeskrivningar i naturligt språk, 1977, ISBN 91-7372-168-9.

No 22 **Jaak Urmi:** A Machine Independent LISP Compiler and its Implications for Ideal Hardware, 1978, ISBN 91-7372-188-3.

No 33 **Tore Risch:** Compilation of Multiple File Queries in a Meta-Database System 1978, ISBN 91-7372-232-4.

No 51 **Erland Jungert:** Synthesizing Database Structures from a User Oriented Data Model, 1980, ISBN 91-7372-387-8.

No 54 **Sture Hägglund:** Contributions to the Development of Methods and Tools for Interactive Design of Applications Software, 1980, ISBN 91-7372-404-1.

No 55 **Pär Emanuelson:** Performance Enhancement in a Well-Structured Pattern Matcher through Partial Evaluation, 1980, ISBN 91-7372-403-3.

No 58 **Bengt Johnsson, Bertil Andersson:** The Human-Computer Interface in Commercial Systems, 1981, ISBN 91-7372-414-9.

No 69 **H. Jan Komorowski:** A Specification of an Abstract Prolog Machine and its Application to Partial Evaluation, 1981, ISBN 91-7372-479-3.

No 71 **René Reboh:** Knowledge Engineering Techniques and Tools for Expert Systems, 1981, ISBN 91-7372-489-0.

No 77 **Östen Oskarsson:** Mechanisms of Modifiability in large Software Systems, 1982, ISBN 91-7372-527-7.

No 94 **Hans Lunell:** Code Generator Writing Systems, 1983, ISBN 91-7372-652-4.

No 97 **Andrzej Lingas:** Advances in Minimum Weight Triangulation, 1983, ISBN 91-7372-660-5.

No 109 **Peter Fritzson:** Towards a Distributed Programming Environment based on Incremental Compilation,1984, ISBN 91-7372-801-2.

No 111 **Erik Tengvald:** The Design of Expert Planning Systems. An Experimental Operations Planning System for Turning, 1984, ISBN 91-7372-805-5.

No 155 **Christos Levcopoulos:** Heuristics for Minimum Decompositions of Polygons, 1987, ISBN 91-7870-133-3.

No 165 **James W. Goodwin:** A Theory and System for Non-Monotonic Reasoning, 1987, ISBN 91-7870-183-X.

No 170 **Zebo Peng:** A Formal Methodology for Automated Synthesis of VLSI Systems, 1987, ISBN 91-7870-225-9.

No 174 **Johan Fagerström:** A Paradigm and System for Design of Distributed Systems, 1988, ISBN 91-7870-301-8.

No 192 **Dimiter Driankov:** Towards a Many Valued Logic of Quantified Belief, 1988, ISBN 91-7870-374-3.

No 213 **Lin Padgham:** Non-Monotonic Inheritance for an Object Oriented Knowledge Base, 1989, ISBN 91-7870-485-5.

No 214 **Tony Larsson:** A Formal Hardware Description and Verification Method, 1989, ISBN 91-7870-517-7.

No 221 **Michael Reinfrank:** Fundamentals and Logical Foundations of Truth Maintenance, 1989, ISBN 91-7870-546-0.

No 239 **Jonas Löwgren:** Knowledge-Based Design Support and Discourse Management in User Interface Management Systems, 1991, ISBN 91-7870-720-X.

No 244 **Henrik Eriksson:** Meta-Tool Support for Knowledge Acquisition, 1991, ISBN 91-7870-746-3.

No 252 **Peter Eklund:** An Epistemic Approach to Interactive Design in Multiple Inheritance Hierarchies,1991, ISBN 91-7870-784-6.

No 258 **Patrick Doherty:** NML3 - A Non-Monotonic Formalism with Explicit Defaults, 1991, ISBN 91-7870-816-8.

No 260 **Nahid Shahmehri:** Generalized Algorithmic Debugging, 1991, ISBN 91-7870-828-1.

No 264 **Nils Dahlbäck:** Representation of Discourse-Cognitive and Computational Aspects, 1992, ISBN 91-7870-850-8.

No 265 **Ulf Nilsson:** Abstract Interpretations and Abstract Machines: Contributions to a Methodology for the Implementation of Logic Programs, 1992, ISBN 91-7870-858-3.

No 270 **Ralph Rönnquist:** Theory and Practice of Tense-bound Object References, 1992, ISBN 91-7870-873-7.

No 273 **Björn Fjellborg:** Pipeline Extraction for VLSI Data Path Synthesis, 1992, ISBN 91-7870-880-X.

No 276 **Staffan Bonnier:** A Formal Basis for Horn Clause Logic with External Polymorphic Functions, 1992, ISBN 91-7870-896-6.

No 277 **Kristian Sandahl:** Developing Knowledge Management Systems with an Active Expert Methodology, 1992, ISBN 91-7870-897-4.

No 281 **Christer Bäckström:** Computational Complexity of Reasoning about Plans, 1992, ISBN 91-7870-979-2.

No 292 **Mats Wirén:** Studies in Incremental Natural Language Analysis, 1992, ISBN 91-7871-027-8.

No 297 **Mariam Kamkar:** Interprocedural Dynamic Slicing with Applications to Debugging and Testing, 1993, ISBN 91-7871-065-0.

No 302 **Tingting Zhang:** A Study in Diagnosis Using Classification and Defaults, 1993, ISBN 91-7871-078-2.

No 312 **Arne Jönsson:** Dialogue Management for Natural Language Interfaces - An Empirical Approach, 1993, ISBN 91-7871-110-X.

No 338 **Simin Nadjm-Tehrani**: Reactive Systems in Physical Environments: Compositional Modelling and Framework for Verification, 1994, ISBN 91-7871-237-8.

No 371 **Bengt Savén:** Business Models for Decision Support and Learning. A Study of Discrete-Event Manufacturing Simulation at Asea/ABB 1968-1993, 1995, ISBN 91-7871-494-X.

No 375 **Ulf Söderman:** Conceptual Modelling of Mode Switching Physical Systems, 1995, ISBN 91-7871-516-4.

No 383 **Andreas Kågedal:** Exploiting Groundness in Logic Programs, 1995, ISBN 91-7871-538-5.

No 396 **George Fodor:** Ontological Control, Description, Identification and Recovery from Problematic Control Situations, 1995, ISBN 91-7871-603-9.

No 413 **Mikael Pettersson:** Compiling Natural Semantics, 1995, ISBN 91-7871-641-1.

No 414 **Xinli Gu:** RT Level Testability Improvement by Testability Analysis and Transformations, 1996, ISBN 91-7871-654-3.

No 416 **Hua Shu:** Distributed Default Reasoning, 1996, ISBN 91-7871-665-9.

No 429 **Jaime Villegas:** Simulation Supported Industrial Training from an Organisational Learning Perspective - Development and Evaluation of the SSIT Method, 1996, ISBN 91-7871-700-0.

No 431 **Peter Jonsson:** Studies in Action Planning: Algorithms and Complexity, 1996, ISBN 91-7871-704-3.

No 437 **Johan Boye:** Directional Types in Logic Programming, 1996, ISBN 91-7871-725-6.

No 439 **Cecilia Sjöberg:** Activities, Voices and Arenas: Participatory Design in Practice, 1996, ISBN 91-7871-728-0.

No 448 **Patrick Lambrix:** Part-Whole Reasoning in Description Logics, 1996, ISBN 91-7871-820-1.

No 452 **Kjell Orsborn:** On Extensible and Object-Relational Database Technology for Finite Element Analysis Applications, 1996, ISBN 91-7871-827-9.

No 459 **Olof Johansson:** Development Environments for Complex Product Models, 1996, ISBN 91-7871-855-4.

No 461 **Lena Strömbäck:** User-Defined Constructions in Unification-Based Formalisms,1997, ISBN 91-7871-857-0.

No 462 **Lars Degerstedt:** Tabulation-based Logic Programming: A Multi-Level View of Query Answering, 1996, ISBN 91-7871-858-9.

No 475 **Fredrik Nilsson:** Strategi och ekonomisk styrning - En studie av hur ekonomiska styrsystem utformas och används efter företagsförvärv, 1997, ISBN 91-7871-914-3.

No 480 **Mikael Lindvall:** An Empirical Study of Requirements-Driven Impact Analysis in Object-Oriented Software Evolution, 1997, ISBN 91-7871-927-5.

No 485 **Göran Forslund**: Opinion-Based Systems: The Cooperative Perspective on Knowledge-Based Decision Support, 1997, ISBN 91-7871-938-0.

No 494 **Martin Sköld**: Active Database Management Systems for Monitoring and Control, 1997, ISBN 91-7219-002-7.

No 495 **Hans Olsén**: Automatic Verification of Petri Nets in a CLP framework, 1997, ISBN 91-7219-011-6.

No 498 **Thomas Drakengren:** Algorithms and Complexity for Temporal and Spatial Formalisms, 1997, ISBN 91-7219-019-1.

No 502 **Jakob Axelsson:** Analysis and Synthesis of Heterogeneous Real-Time Systems, 1997, ISBN 91-7219-035-3.

No 503 **Johan Ringström:** Compiler Generation for Data-Parallel Programming Langugaes from Two-Level Semantics Specifications, 1997, ISBN 91-7219-045-0.

No 512 **Anna Moberg:** Närhet och distans - Studier av kommunikationsmmönster i satellitkontor och flexibla kontor, 1997, ISBN 91-7219-119-8.

No 520 **Mikael Ronström:** Design and Modelling of a Parallel Data Server for Telecom Applications, 1998, ISBN 91-7219-169-4.

No 522 **Niclas Ohlsson:** Towards Effective Fault Prevention - An Empirical Study in Software Engineering, 1998, ISBN 91-7219-176-7.

No 526 **Joachim Karlsson:** A Systematic Approach for Prioritizing Software Requirements, 1998, ISBN 91-7219-184-8.

No 530 **Henrik Nilsson:** Declarative Debugging for Lazy Functional Languages, 1998, ISBN 91-7219-197-x.

No 555 **Jonas Hallberg:** Timing Issues in High-Level Synthesis,1998, ISBN 91-7219-369-7.

No 561 **Ling Lin:** Management of 1-D Sequence Data - From Discrete to Continuous, 1999, ISBN 91-7219-402-2.

No 563 **Eva L Ragnemalm:** Student Modelling based on Collaborative Dialogue with a Learning Companion, 1999, ISBN 91-7219-412-X.

No 567 **Jörgen Lindström:** Does Distance matter? On geographical dispersion in organisations, 1999, ISBN 91-7219-439-1.

No 582 **Vanja Josifovski:** Design, Implementation and

Evaluation of a Distributed Mediator System for Data Integration, 1999, ISBN 91-7219-482-0.

No 589 **Rita Kovordányi**: Modeling and Simulating Inhibitory Mechanisms in Mental Image Reinterpretation - Towards Cooperative Human-Computer Creativity, 1999, ISBN 91-7219-506-1.

No 592 **Mikael Ericsson:** Supporting the Use of Design Knowledge - An Assessment of Commenting Agents, 1999, ISBN 91-7219-532-0.

No 593 **Lars Karlsson:** Actions, Interactions and Narratives, 1999, ISBN 91-7219-534-7.

No 594 **C. G. Mikael Johansson:** Social and Organizational Aspects of Requirements Engineering Methods - A practice-oriented approach, 1999, ISBN 91-7219-541-X.

No 595 **Jörgen Hansson:** Value-Driven Multi-Class Overload Management in Real-Time Database Systems, 1999, ISBN 91-7219-542-8.

No 596 **Niklas Hallberg:** Incorporating User Values in the Design of Information Systems and Services in the Public Sector: A Methods Approach, 1999, ISBN 91-7219-543-6.

No 597 **Vivian Vimarlund:** An Economic Perspective on the Analysis of Impacts of Information Technology: From Case Studies in Health-Care towards General Models and Theories, 1999, ISBN 91-7219-544-4.

No 598 **Johan Jenvald:** Methods and Tools in Computer-Supported Taskforce Training, 1999, ISBN 91-7219-547-9.

No 607 **Magnus Merkel:** Understanding and enhancing translation by parallel text processing, 1999, ISBN 91-7219-614-9.

No 611 **Silvia Coradeschi:** Anchoring symbols to sensory data, 1999, ISBN 91-7219-623-8.

No 613 **Man Lin:** Analysis and Synthesis of Reactive Systems: A Generic Layered Architecture Perspective, 1999, ISBN 91-7219-630-0.

No 618 **Jimmy Tjäder:** Systemimplementering i praktiken - En studie av logiker i fyra projekt, 1999, ISBN 91-7219-657-2.

No 627 **Vadim Engelson:** Tools for Design, Interactive Simulation, and Visualization of Object-Oriented Models in Scientific Computing, 2000, ISBN 91-7219-709-9.

No 637 **Esa Falkenroth:** Database Technology for Control and Simulation, 2000, ISBN 91-7219-766-8.

No 639 **Per-Arne Persson:** Bringing Power and Knowledge Together: Information Systems Design for Autonomy and Control in Command Work, 2000, ISBN 91-7219-796-X.

No 660 **Erik Larsson:** An Integrated System-Level Design for Testability Methodology, 2000, ISBN 91-7219-890-7.

No 688 **Marcus Bjäreland:** Model-based Execution Monitoring, 2001, ISBN 91-7373-016-5.

No 689 **Joakim Gustafsson:** Extending Temporal Action Logic, 2001, ISBN 91-7373-017-3.

No 720 **Carl-Johan Petri:** Organizational Information Provision - Managing Mandatory and Discretionary Use of Information Technology, 2001, ISBN-91-7373-126-9.

No 724 **Paul Scerri:** Designing Agents for Systems with Adjustable Autonomy, 2001, ISBN 91 7373 207 9.

No 725 **Tim Heyer**: Semantic Inspection of Software Artifacts: From Theory to Practice, 2001, ISBN 91 7373 208 7.

No 726 **Pär Carlshamre:** A Usability Perspective on Requirements Engineering - From Methodology to Product Development, 2001, ISBN 91 7373 212 5.

No 732 **Juha Takkinen:** From Information Management to Task Management in Electronic Mail, 2002, ISBN 91 7373 258 3.

No 745 **Johan Åberg:** Live Help Systems: An Approach to Intelligent Help for Web Information Systems, 2002, ISBN 91-7373-311-3.

No 746 **Rego Granlund:** Monitoring Distributed Teamwork Training, 2002, ISBN 91-7373-312-1.

No 757 **Henrik André-Jönsson:** Indexing Strategies for Time Series Data, 2002, ISBN 917373-346-6.

No 747 **Anneli Hagdahl:** Development of IT-suppor-ted Inter-organisational Collaboration - A Case Study in the Swedish Public Sector, 2002, ISBN 91-7373-314-8.

No 749 **Sofie Pilemalm:** Information Technology for Non-Profit Organisations - Extended Participatory Design of an Information System for Trade Union Shop Stewards, 2002, ISBN 91-7373-318-0.

No 765 **Stefan Holmlid:** Adapting users: Towards a theory of use quality, 2002, ISBN 91-7373-397-0.

No 771 **Magnus Morin:** Multimedia Representations of Distributed Tactical Operations, 2002, ISBN 91-7373-421-7.

No 772 **Pawel Pietrzak:** A Type-Based Framework for Locating Errors in Constraint Logic Programs, 2002, ISBN 91-7373-422-5.

No 758 **Erik Berglund:** Library Communication Among Programmers Worldwide, 2002, ISBN 91-7373-349-0.

No 774 **Choong-ho Yi:** Modelling Object-Oriented Dynamic Systems Using a Logic-Based Framework, 2002, ISBN 91-7373-424-1.

No 779 **Mathias Broxvall:** A Study in the Computational Complexity of Temporal Reasoning, 2002, ISBN 91-7373-440-3.

No 793 **Asmus Pandikow:** A Generic Principle for Enabling Interoperability of Structured and Object-Oriented Analysis and Design Tools, 2002, ISBN 91-7373-479-9.

No 785 **Lars Hult:** Publika Informationstjänster. En studie av den Internetbaserade encyklopedins bruksegenskaper, 2003, ISBN 91-7373-461-6.

No 800 **Lars Taxén:** A Framework for the Coordination of Complex Systems´ Development, 2003, ISBN 91-7373-604-X

No 808 **Klas Gäre:** Tre perspektiv på förväntningar och förändringar i samband med införande av informationsystem, 2003, ISBN 91-7373-618-X.

No 821 **Mikael Kindborg:** Concurrent Comics - programming of social agents by children, 2003, ISBN 91-7373-651-1.

No 823 **Christina Ölvingson:** On Development of Information Systems with GIS Functionality in Public Health Informatics: A Requirements Engineering Approach, 2003, ISBN 91-7373-656-2.

No 828 **Tobias Ritzau:** Memory Efficient Hard Real-Time Garbage Collection, 2003, ISBN 91-7373-666-X.

No 833 **Paul Pop:** Analysis and Synthesis of Communication-Intensive Heterogeneous Real-Time Systems, 2003, ISBN 91-7373-683-X.

No 852 **Johan Moe:** Observing the Dynamic Behaviour of Large Distributed Systems to Improve Development and Testing - An Emperical Study in Software Engineering, 2003, ISBN 91-7373-779-8.

No 867 **Erik Herzog:** An Approach to Systems Engineering Tool Data Representation and Exchange, 2004, ISBN 91-7373-929-4.

No 872 **Aseel Berglund:** Augmenting the Remote Control: Studies in Complex Information Navigation for Digital TV, 2004, ISBN 91-7373-940-5.

No 869 **Jo Skåmedal:** Telecommuting's Implications on Travel and Travel Patterns, 2004, ISBN 91-7373-935-9.

No 870 **Linda Askenäs:** The Roles of IT - Studies of Organising when Implementing and Using Enterprise Systems, 2004, ISBN 91-7373-936-7.

No 874 **Annika Flycht-Eriksson:** Design and Use of Ontologies in Information-Providing Dialogue Systems, 2004, ISBN 91-7373-947-2.

No 873 **Peter Bunus:** Debugging Techniques for Equation-Based Languages, 2004, ISBN 91-7373-941-3.

No 876 **Jonas Mellin:** Resource-Predictable and Efficient Monitoring of Events, 2004, ISBN 91-7373-956-1.

No 883 **Magnus Bång:** Computing at the Speed of Paper: Ubiquitous Computing Environments for Healthcare Professionals, 2004, ISBN 91-7373-971-5

No 882 **Robert Eklund:** Disfluency in Swedish human-human and human-machine travel booking dialogues, 2004. ISBN 91-7373-966-9.

No 887 **Anders Lindström:** English and other Foreign Linquistic Elements in Spoken Swedish. Studies of Productive Processes and their Modelling using Finite-State Tools, 2004, ISBN 91-7373-981-2.

No 889 **Zhiping Wang:** Capacity-Constrained Production-inventory systems - Modellling and Analysis in both a traditional and an e-business context, 2004, ISBN 91-85295-08-6.

No 893 **Pernilla Qvarfordt:** Eyes on Multimodal Interaction, 2004, ISBN 91-85295-30-2.

No 910 **Magnus Kald:** In the Borderland between Strategy and Management Control - Theoretical Framework and Empirical Evidence, 2004, ISBN 91-85295-82-5.

No 918 **Jonas Lundberg:** Shaping Electronic News: Genre Perspectives on Interaction Design, 2004, ISBN 91-85297-14-3.

No 900 **Mattias Arvola:** Shades of use: The dynamics of interaction design for sociable use, 2004, ISBN 91-85295-42-6.

No 920 **Luis Alejandro Cortés:** Verification and Scheduling Techniques for Real-Time Embedded Systems, 2004, ISBN 91-85297-21-6.

No 929 **Diana Szentivanyi:** Performance Studies of Fault-Tolerant Middleware, 2005, ISBN 91-85297-58-5.

No 933 **Mikael Cäker:** Management Accounting as Constructing and Opposing Customer Focus: Three Case Studies on Management Accounting and Customer Relations, 2005, ISBN 91-85297-64-X.

No 937 **Jonas Kvarnström:** TALplanner and Other Extensions to Temporal Action Logic, 2005, ISBN 91-85297-75-5.

No 938 **Bourhane Kadmiry:** Fuzzy Gain-Scheduled Visual Servoing for Unmanned Helicopter, 2005, ISBN 91-85297-76-3.

No 945 **Gert Jervan:** Hybrid Built-In Self-Test and Test Generation Techniques for Digital Systems, 2005, ISBN: 91-85297-97-6.

No 946 **Anders Arpteg:** Intelligent Semi-Structured Information Extraction, 2005, ISBN 91-85297-98-4.

No 947 **Ola Angelsmark:** Constructing Algorithms for Constraint Satisfaction and Related Problems - Methods and Applications, 2005, ISBN 91-85297-99-2.

No 963 **Calin Curescu:** Utility-based Optimisation of Resource Allocation for Wireless Networks, 2005. ISBN 91-85457-07-8.

No 972 **Björn Johansson:** Joint Control in Dynamic Situations, 2005, ISBN 91-85457-31-0.

No 974 **Dan Lawesson:** An Approach to Diagnosability Analysis for Interacting Finite State Systems, 2005, ISBN 91-85457-39-6.

No 979 **Claudiu Duma:** Security and Trust Mechanisms for Groups in Distributed Services, 2005, ISBN 91-85457-54-X.

No 983 **Sorin Manolache:** Analysis and Optimisation of Real-Time Systems with Stochastic Behaviour, 2005, ISBN 91-85457-60-4.

No 986 **Yuxiao Zhao:** Standards-Based Application Integration for Business-to-Business Communications, 2005, ISBN 91-85457-66-3.

No 1004 **Patrik Haslum:** Admissible Heuristics for Automated Planning, 2006, ISBN 91-85497-28-2.

No 1005 **Aleksandra Tešanovic:** Developing Reusable and Reconfigurable Real-Time Software using Aspects and Components, 2006, ISBN 91-85497-29-0.

No 1008 **David Dinka:** Role, Identity and Work: Extending the design and development agenda, 2006, ISBN 91-85497-42-8.

No 1009 **Iakov Nakhimovski:** Contributions to the Modeling and Simulation of Mechanical Systems with Detailed Contact Analysis, 2006, ISBN 91-85497-43-X.

No 1013 **Wilhelm Dahllöf:** Exact Algorithms for Exact Satisfiability Problems, 2006, ISBN 91-85523-97-6.

No 1016 **Levon Saldamli:** PDEModelica - A High-Level Language for Modeling with Partial Differential Equations, 2006, ISBN 91-85523-84-4.

No 1017 **Daniel Karlsson:** Verification of Component-based Embedded System Designs, 2006, ISBN 91-85523-79-8.

No 1018 **Ioan Chisalita:** Communication and Networking Techniques for Traffic Safety Systems, 2006, ISBN 91-85523-77-1.

No 1019 **Tarja Susi:** The Puzzle of Social Activity - The Significance of Tools in Cognition and Cooperation, 2006, ISBN 91-85523-71-2.

No 1021 **Andrzej Bednarski:** Integrated Optimal Code Generation for Digital Signal Processors, 2006, ISBN 91-85523-69-0.

No 1022 **Peter Aronsson:** Automatic Parallelization of Equation-Based Simulation Programs, 2006, ISBN 91-85523-68-2.

No 1023 **Sonia Sangari:** Some Visual Correlates to Focal Accent in Swedish, 2006, ISBN 91-85523-67-4.

No 1030 **Robert Nilsson:** A Mutation-based Framework for Automated Testing of Timeliness, 2006, ISBN 91-85523-35-6.

No 1034 **Jon Edvardsson:** Techniques for Automatic Generation of Tests from Programs and Specifications, 2006, ISBN 91-85523-31-3.

No 1035 **Vaida Jakoniene:** Integration of Biological Data, 2006, ISBN 91-85523-28-3.

No 1045 **Genevieve Gorrell:** Generalized Hebbian Algorithms for Dimensionality Reduction in Natural Language Processing, 2006, ISBN 91-85643-88-2.

No 1051 **Yu-Hsing Huang:** Having a New Pair of Glasses - Applying Systemic Accident Models on Road Safety, 2006, ISBN 91-85643-64-5.

No 1054 **Åsa Hedenskog:** Perceive those things which cannot be seen - A Cognitive Systems Engineering perspective on requirements management, 2006, ISBN 91-85643-57-2.

No 1061 **Cécile Åberg:** An Evaluation Platform for Semantic Web Technology, 2007, ISBN 91-85643-31-9.

No 1073 **Mats Grindal:** Handling Combinatorial Explosion in Software Testing, 2007, ISBN 978-91-85715-74-9.

No 1075 **Almut Herzog:** Usable Security Policies for Runtime Environments, 2007, ISBN 978-91-85715-65-7.

No 1079 **Magnus Wahlström:** Algorithms, measures, and upper bounds for satisfiability and related problems, 2007, ISBN 978-91-85715-55-8.

No 1083 **Jesper Andersson:** Dynamic Software Architectures, 2007, ISBN 978-91-85715-46-6.

No 1086 **Ulf Johansson:** Obtaining Accurate and Comprehensible Data Mining Models - An Evolutionary Approach, 2007, ISBN 978-91-85715-34-3.

No 1089 **Traian Pop:** Analysis and Optimisation of Distributed Embedded Systems with Heterogeneous Scheduling Policies, 2007, ISBN 978-91-85715-27-5.

No 1091 **Gustav Nordh:** Complexity Dichotomies for CSP-related Problems, 2007, ISBN 978-91-85715-20-6.

No 1106 **Per Ola Kristensson:** Discrete and Continuous Shape Writing for Text Entry and Control, 2007, ISBN 978-91-85831-77-7.

No 1110 **He Tan:** Aligning Biomedical Ontologies, 2007, ISBN 978-91-85831-56-2.

Linköping Studies in Information Science

No 1 **Karin Axelsson:** Metodisk systemstrukturering- att skapa samstämmighet mellan informa-tionssystemarkitektur och verksamhet, 1998. ISBN-9172-19-296-8.

No 2 **Stefan Cronholm:** Metodverktyg och användbarhet - en studie av datorstödd metodbaserad systemutveckling, 1998. ISBN-9172-19-299-2.

No 3 **Anders Avdic:** Användare och utvecklare - om anveckling med kalkylprogram, 1999. ISBN-91-7219-606-8.

No 4 **Owen Eriksson:** Kommunikationskvalitet hos informationssystem och affärsprocesser, 2000. ISBN 91-7219-811-7.

No 5 **Mikael Lind:** Från system till process - kriterier för processbestämning vid verksamhetsanalys, 2001, ISBN 91-7373-067-X

No 6 **Ulf Melin:** Koordination och informationssystem i företag och nätverk, 2002, ISBN 91-7373-278-8.

No 7 **Pär J. Ågerfalk:** Information Systems Actability - Understanding Information Technology as a Tool for Business Action and Communication, 2003, ISBN 91-7373-628-7.

No 8 **Ulf Seigerroth:** Att förstå och förändra systemutvecklingsverksamheter - en taxonomi för metautveckling, 2003, ISBN91-7373-736-4.

No 9 **Karin Hedström:** Spår av datoriseringens värden - Effekter av IT i äldreomsorg, 2004, ISBN 91-7373-963-4.

No 10 **Ewa Braf:** Knowledge Demanded for Action - Studies on Knowledge Mediation in Organisations, 2004, ISBN 91-85295-47-7.

No 11 **Fredrik Karlsson:** Method Configuration - method and computerized tool support, 2005, ISBN 91-85297-48-8.

No 12 **Malin Nordström:** Styrbar systemförvaltning - Att organisera systemförvaltningsverksamhet med hjälp av effektiva förvaltningsobjekt, 2005, ISBN 91-85297-60-7.

No 13 **Stefan Holgersson:** Yrke: POLIS - Yrkeskunskap, motivation, IT-system och andra förutsättningar för polisarbete, 2005, ISBN 91-85299-43-X.

No 14 **Benneth Christiansson, Marie-Therese Christiansson:** Mötet mellan process och komponent - mot ett ramverk för en verksamhetsnära kravspecifikation vid anskaffning av komponentbaserade informationssystem, 2006, ISBN 91-85643-22-X.