KNOWLEDGE MEDIA
KMi
INSTITUTE

Probabilistic Methods for Data Integration in a Multi-Agent Query Answering System

Miklós Nagy

Supervisors: Maria Vargas-Vera, Enrico Motta

KMI-TR-08-05

October, 2008

http://kmi.open.ac.uk/publications/pdf/kmi-tr-08-05.pdf

The Open University

Contents

1 Abstract

This report describes the progress that has been achieved during the second year (full time equivalent 1 July 2006 - 1 July 2008) of our Ph.D. research. All the work has been built upon the achievements of the first year and confirmed that the original research objectives were correctly identified at the beginning of the research. We have successfully participated in the Ontology Mapping Evaluation Initiative 2006 and 2007 (2008 ongoing activity), which provided a qualitative comparison of our and other ontology mapping systems. Further it created a possibility to identify our future research work that needs to be carried out in order to achieve our original research objectives that were set out in the formal Ph.D. research proposal. The organization of this report corresponds to original research proposal. The last section describes the thesis outline.

2 Introduction

The standard and most acceptable view of the Semantic Web assumes that we will not have single consensus ontology for a given domain, but many, each with its own base of users and applications. Our vision is that today's popular search engines will evolve into question answering systems that will be able to

1. Answer concrete user queries utilizing semantic distributed knowledge (ontologies).
2. Integrate information.

that is represented on the Semantic Web.

The success of such query answering application that can make use of these distributed ontologies will depend on how correctly the system can create mapping and discover conceptual links between the content of these knowledge sources. Most of the current mapping solutions and research require human experts for validating such mappings. Increasing number of prototype applications and research is focusing on utilizing probabilistic information to address the problem of finding approximate ontology mappings when perfect mappings cannot be found. It has been suggested that background knowledge e.g. can be considered as a source of a statistic sample to train machine-learning algorithms, which in turn can provide better mappings for real data sets [1]. We believe that ontology mapping can considerably be improved by using the background knowledge to calculate subjective probabilities (beliefs) about the correctness of the mapping. In my Ph.D. research I mainly focus on Dempster Shafer theory which I believe to have some remarkable advantages such as the ability to represent missing information (ignorance) over the traditional Bayesian solutions.

3 Similarity mapping algorithms and measures in a distributed environment

In the question answering context we have investigated that if the knowledge or information that is necessary to built up a mapping is distributed among domain specific agents (each agent has information about a set of closely related concepts or group of concepts) then

1. How the correctness of the mapping will be affected?
2. What kind of similarity measures and distributed algorithms can improve the results that have been achieved with the current solutions?

In our prototype we have created a number of agents, which hold distributed knowledge of the specific domain. We have found that the answer quality point of view the correctness of the mapping is not dependent on the knowledge distribution if the query is distributed correctly. This confirmed our assumption that distributed knowledge does not decrease the quality of the mapping. However we have to assume that our query is distributed correctly because the investigation of such distribution algorithm was out of the scope of my Ph.D. research. Further we have identified that local background knowledge is a key factor in producing correct similarities. We have investigated using both custom developed dictionary and WordNet as a local background knowledge. Our conclusion was that using WordNet produces more uncertain alternatives than using well-established custom specific dictionary. For our further stages of our research we need considerable amount of uncertain mapping candidates hence it is advisable to keep WordNet as local background knowledge source however we need to enrich the standard WordNet dictionary with some domain specific information. Nevertheless the utilization of multi lingual domains specific background knowledge is considered as a further research direction for our system.

4 Role of distributed local knowledge in ontology mapping

Concerning the knowledge management perspective of our research it is based on the Distributed Knowledge Management (DKM) approach [2], in which subjective and social aspects of the real world are seriously taken into account. Compared to the traditional Knowledge Management (KM) view of creating, codifying and disseminating knowledge as single, supposedly shared and objective classification we believe that Distributed Knowledge Management is the viable alternative to the concept of an existing "centralized or common knowledge" in the context of question answering on the WWW. The concept of Semantic Web is also based on the distributed knowledge idea, which is represented by different ontologies. However in the context of ontology mapping and information integration the current state of the art approaches does not reflect fully these ideas since as our literature review pointed out most of the mapping approaches were based on the centralized or common knowledge approach.
The reason we investigated the distributed approach further in our research is because in complex environments like the question answering the knowledge is:

1. Locally defined and based on the entities perspective i.e. subjective.
2. Exchangeable between the local perspectives.

Our basic argument in my Ph.D. research that knowledge cannot be viewed as a simple conceptualization of the world, but it has to represent some degree of interpretation. Such interpretation depends on the context of the entities involved in the process. However this implies that the resulting mapping will be based on some sort of subjective belief, which needs to be handled correctly. Experiments with real word ontologies that have been carried out in the Ontology Alignment Initiative 2006 and 2007 clearly confirmed that our theory is correct. To illustrate the problem we took two concepts from real word ontologies

- Ontology 1: MScThesis

- Ontology 2: MasterThesis

Current state of the art solutions are based on structure or semantic based techniques to determine that these concepts are the same since no similarity measure can provide a qualitative comparison with considerable certainly. For this problem we have proposed and implemented a structure based method, which considers the local context of the ontology by incorporating the local context and its synonyms from WordNet into the original comparison e.g. an MScThesis is a **report** made by a **student.** Our conclusion was that our original idea was correct however we need to consider that any examination of the local context should involve dividing concepts or properties into syllables if possible otherwise it cannot really be associated with its context e.g. MScThesis does not exist in WordNet but MSc and Thesis separately does. Such word division needs to be incorporated into our algorithms.

5 Incorporating trust in the mapping process

In the Semantic Web, all kinds of information are expressed on a single information model framework, which allows us to connect different kinds of information from different sources and use them as a huge distributed database. In the present Semantic Web there is no mechanism for evaluating the trust of the particular sources. Theoretically everyone can freely write and publish information on their Web page or in a database, which involves the possibility of incorporating incorrect information.

Trust and trustworthiness can be defined as the degree to which one considers an assertion to be true for a given context, where trust is often used to interpret a very high degree of confidence. Therefore the problem of trustworthiness on the Semantic Web can emerge in different levels e.g. trust in the available semantic description of data or the available information after processing such data. Trust can be represented differently depending of its nature and can be based on the followings:

1. Reputation: use rating systems and web of trust mechanisms
2. Context: use background information about the information provider
3. Content: use information content itself, together with the related information content published by other information providers.

Dominantly the existing approaches that address the problem of the trustworthiness of the available data on the Semantic Web are reputation and context based e.g. using digital signatures that would state who the publisher of the ontology is. In the context of question answering the issue of trust in the different sources can significantly affect the overall system acceptance by the users thus the practical usability of the system. There are several ways to define trust as the quantified belief by a trustor within a specified context. Quantification reflects that a trustor can have various degrees of trust, which could be expressed as a numerical range or as a simple semantic classification. However a trust level for one context doesn't normally apply to a different context hence the attributes of trust is dependent on the trust context. In many practical situations, there might not be any past experience for a specific trustee or context on which to base a trust evaluation. Thus the evaluation might have to depend on trust evaluation from a different context.

Therefore some kind of trust evaluation should exist between agents and sources.

The research questions that had been investigated during this phase were:

1. What factors had to be considered when expressing such initial trust?
2. How to express these factors in a numerical way?
3. How to maintain consistent trust factors between the agents and sources?
4. How these factors can be effectively incorporated into a probabilistic framework?

Based on our experiments with the Ontology Evaluation Initiative 2006 and 2007 benchmarks we have investigated cases where the belief combination produced incorrect result even thought before the combination a correct mapping could have been derived for the particular case based on individual beliefs. The problem occurred when the different agents' similarity assessment produces conflicting beliefs over the correctness of a particular mapping. A conflict between two beliefs in DS theory can be interpreted qualitatively as one source strongly supports one hypothesis and the other strongly supports another hypothesis, and the two hypotheses are not compatible. In this scenario applying Dempster’s combination rule to conflicting beliefs can lead to an almost impossible choice with a very low degree of belief, which will result in the most possible outcome with a very high degree of belief [3, 4]. This combination rule strongly emphasizes the agreement between multiple sources and ignores all the conflicting evidence through a normalization factor.
In our ontology mapping framework we propose that the belief functions are considered as a method to model an agent’s beliefs, therefore the belief function defined by an agent can also be viewed as a way of expressing the agent’s preferences over choices, with respect to masses assigned to different hypotheses. The larger the mass assigned to a hypothesis is the more preferred the hypothesis would be. In this context the problem is how do we handle the agent's conflicting individual preferences that need to be aggregated in order to form a collective preference.

6 Conflicting evidence and the fuzzy voting model

In ontology mapping the conflicting results of the different similarity measures can be resolved if the mapping algorithm can produce an agreed solution, even though the individual opinions about the available alternatives may vary. We propose a solution for reaching this agreement by evaluating trust between the proposed similarities through voting which is a general method of reconciling differences. Voting is a mechanism where the opinions from a set of votes are evaluated in order to select the alternatives that best represent the collective preferences. Unfortunately deriving binary trust like trustful or not trustful from the difference of belief functions is not so straightforward since the different voters express their opinion as subjective probability over the similarities. For a particular mapping this always involves a certain degree of vagueness hence the threshold between the trust and no trust cannot be set definitely for all cases that can occur during the process. Our argument is that the trust membership value, which is expressed by different voters, can be modelled properly by using fuzzy representation as depicted on Figure 1.

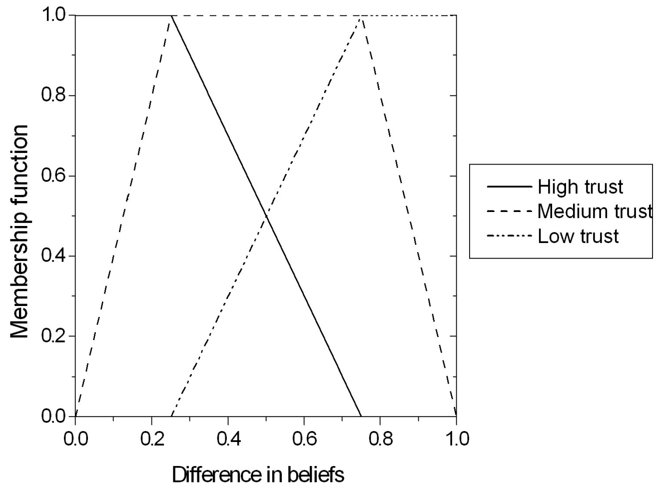

Figure 1

In fuzzy logic the membership function μ(x) is defined on the universe of discourse U and represents a particular input value as a member of the fuzzy set i.e. μ(x) is a curve that defines how each point in the U is mapped to a membership value (or degree of membership) between 0 and 1. For representing trust in beliefs over similarities we have defined three overlapping trapezoidal membership functions, which represents

- High trust
- Medium trust
- Low trust

in the beliefs over concept and property similarities in our ontology mapping system.

7 Converting similarity measures into belief masses

In this phase of the research we were trying to answer a single but very complex question:

1. Where the belief masses will come from i.e. how can we use similarity measures to back our hypothesis and be able to make subjective judgment about the probabilities?

The subjective judgment or probability is the agent's actual judgment, normally representing what a human expert's judgment would be, in view of his information to date and of his sense of other people's information, even if the particular judgment is not shared by the other experts.

The question made sense because our similarity algorithms can provide numerical values between certain types of concept name or property matches however these numbers need to be translated into a quantitative form of belief.

This was not a straightforward question and we have reviewed several studies that tried to investigate how probability can be assessed from similarities [6,7,8].

As the literature review showed researchers have begun investigating the Bayesian theorem as a probabilistic framework for handling uncertainty in the context of ontology mapping. The reason why we chose to investigate Dempster-Shafer theory in our multi agent framework for ontology mapping is because we believed that our approach would prove to be more effective. We based our belief on the fact that according to the behavioral theory the human reasoning about probability rarely follows Bayes's Theorem [9]. In our research we established probability statements under this concept to represent the degree of rational belief that the agent holds about the likelihood of correctness of the mapping. Investigating this kind of subjective concept of probability enabled us to assess the usability of probability statements about the mappings even if these are non-repeatable events, and provides a mechanism for formulating and understanding practical beliefs about probabilities.
We wanted to investigate the use of heuristics in the estimations of likelihood of uncertain events however we decided that it should be out of the scope of our research because it can make our system too dependent on a particular context. Applying our algorithm in the context of the Ontology Alignment Initiative 2006 which contains ontologies from different domains, we have found that any heuristics, of which representativeness and availability can make estimation of probabilities computationally untraceable, often results in biases (violations of the probability axioms) when applied in different domain.

We have found that Dempster-Shafer theory has started to be investigated by an another research groups [10, 12] in the context of ontology mapping, however their method differ from ours because

- They drop the closed word assumption [11] thus it is not necessary to normalize the state space, which resolves the problem of combining contradictory similarity assessments by assuming that the real mapping cannot be determined between the existing concepts or properties.
- They do not use any background knowledge for assessing the particular mapping but combine beliefs over the correctness of the mapping between the existing source concept/property and the full set of existing target concepts/properties.

In their approach they do not consider any optimisation of the belief combination and do not report any performance issues concerning the mapping process. Moreover they have not participated with their system in the Ontology Evaluation Initiative thus we cannot really make a comparison between the two approaches at the moment. However it is worth to keep an eye on their research work since it is slightly overlapping but mainly complimentary to our research.

8 Efficient uncertain reasoning for ontology mapping

Ontology mapping algorithms can use different kind of background knowledge e.g. WordNet, Semantic Web in order to provide better ontology mappings. From the reasoning point of view this additional knowledge increases the

number of variables in the system. As a consequence these additional variables increase the computational complexity of the reasoning process, which is important if the mapping system need to provide response in real time.
Using the Dempster Shafer theory for managing the uncertain aspects of ontology mapping can improve the mapping precision but certain limitation needs to be introduced in order to keep the response time of the mapping algorithm within acceptable limits. One of these limitations as stated in (Nagy, Vargas-Vera, &Motta 2006a) is to limit the number of additional variables that stem from the utilisation of the background knowledge in order to avoid the problem of large problem space, which grows exponentially as the number of variable increases. However the introduction of any limitation can negatively affect the mapping precision because it implies that some background knowledge cannot be completely utilised. In order to make the belief combination and reasoning feasible for large number of variables we need to apply optimisation during the computation. Junction tree based optimisations are ideal for distributed architecture because the problem space can be split up into several portions. In practice however a near optimal junction tree does not guarantee that the split will result in equal portions in terms of variable size which hinders the practical distributed implementation because the computational load cannot be distributed equally (see Figure 2).

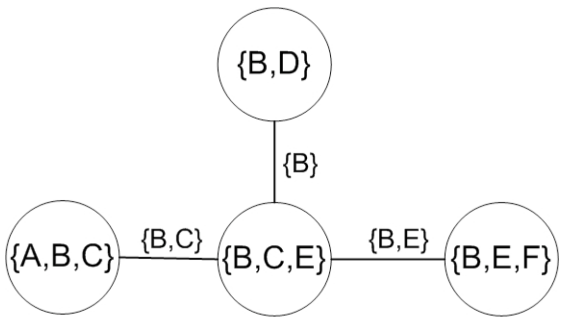

Figure 2

In order to create a junction tree we propose using genetic algorithm, which works under limited time constraint. This is especially important because in our multi agent ontology-mapping framework we need to create mappings in real time and interact with the users if necessary. The main advantages of using genetic algorithm for finding a good junction tree are as follows:

1. Evolution can be distributed between agents
2. Once a time limit is reached the most fittest candidate can be selected from the population

Our objective is to create an algorithm, which finds a good junction tree that can be distributed into our different agents so the reasoning process for the mapping selection can be carried out within acceptable time frame. We have tested our approach to measure how much time is necessary in order to create a junction tree. For this test we have defined 25 variables which represent the possible solutions in the ontologies i.e. concepts or properties. Using Dempster Shafer combination 25 distinct variables would result in a possible problem space with the size 225 = 33, 554, 432. For our experiments we have defined two sets of experiment where we consider using 5 agents, which is represented by a chromosome with 5 genes. We have simulated 1000 junction creation cycles where we have measured the number of evolutions that was necessary to evolve a chromosome, which describes a junction tree for the agents. In every cycle we have evolved the population till the point where the evolution was able to produce a fitter individual compared to the previous generation. The maximum number of evolution was set to 100 because this is the limit that represents a time constraint that can be used in order to make our algorithm responsive from the user point of view. The initial population size in every cycle was set to 50 and the initial populations were generated randomly. In each cycle we have simulated a random number of similarity assignments hence evidences with a random number of variables in each set. The result of our first experiment is depicted on Figure 3

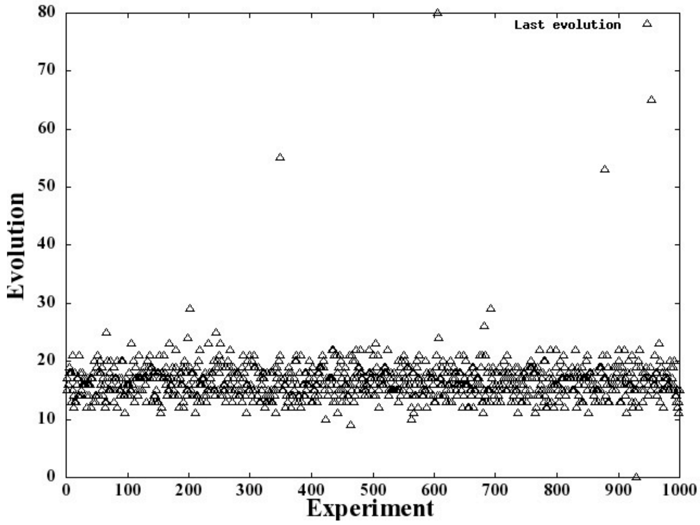

Figure 3

Results have shown that the average number of necessary evolutions to create the junction tree was relatively stable for all tests, which are encouraging considering our original objective that our process should operate under time constraint. Our solution is especially suitable for a multi agent ontology mapping where the mapping system should interact with the users and where the number of possible variables for the mapping selection is large.

9 Implementation

As part of the research objectives we develop a prototype multi agent system, which serves as a test bed for our experiments and allows us to evaluate and prove our theoretical work with real experiments. During the first year we have created the base of our multi agent system using the Jade agent framework. During the second year we have slightly extended out prototype with a Web interface, which will allow us in the future to deploy the system into a Web server. At the moment this web interface can be reached from the Internet on demand however it requires special server side settings since it is deployed on a laptop that obtains IP address dynamically. During the second year no significant development has been made on the multi agent system because participation in the Ontology Alignment Initiative 2006 and 2007 required the mapping algorithm as a standalone process which had to be implemented using a java api that was provided by the organizers. Originally our plan was to create our multi agent mapping system using ontologies from the material science domain however due to practical reasons we choose to continue with the ontologies provided by the Ontology Evaluation Initiative. This is especially useful since the alignment contest has several tracks one of them an “external evaluation” which allows the organizers to run test without the participants. Therefore we have integrated our multi agent prototype with the Ontology Evaluation Initiative api and we were able to participate in other tracks other than the benchmarks. The existing prototype architecture is depicted on Figure 4.

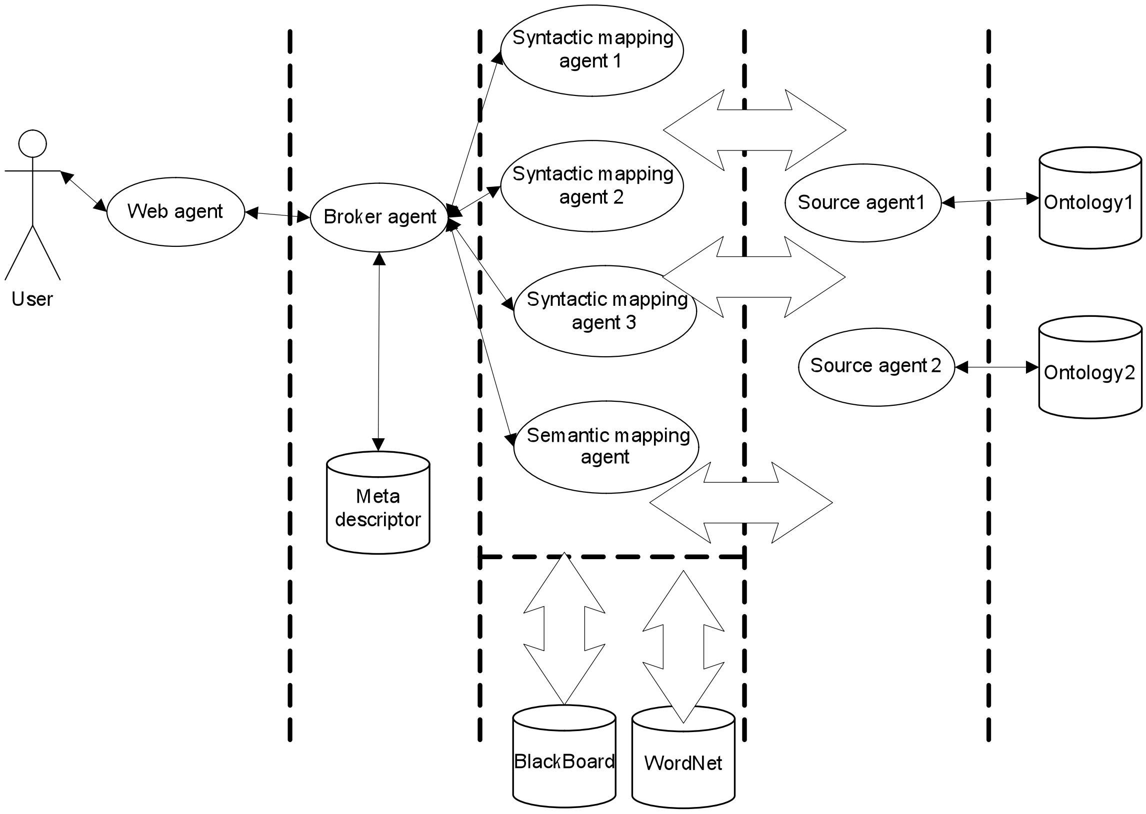

Figure 4

10 Comparison with other approaches

Based on our participation on the OAEI 2006 and 2007 we can continuously assess qualitative comparison between our and other systems. This year we have targeted participation in as much tracks as possible. Considering the track coverage we have performed really well this year compared to other systems. In total 17 systems have participated last year and their test coverage has been as follows:

Software	benchmark	anatomy	directory	food	environment	library	conference	Σ
Falcon-AO v0.7	•	•	•	•	•	•	•	7
DSSim	•	•	•	•	•	•		6
ASMOV	•	•	•				•	4
Lily	•	•	•				•	4
Prior+	•	•	•	•				4
RiMOM	•	•	•	•				4
X-SOM	•	•	•	•				4
OLA2	•		•				•	3
OntoDNA	•		•				•	3
SAMBO	•	•						2
SEMA	•						•	2
TaxoMap	•	•						2
AgreementMaker		•						1
AOAS		•						1
OWL-CM	•							1
SCARLET				•				1
Silas						•		1

Table 1 Tracks per system

Table 1 describes in which tracks the different systems have participated in the OAEI 2007 ordered by the total number of tracks (last column).

The benchmark track is the most popular since it contains the mappings as well and it is possible to compare our results with the reference mappings. Considering the precision and recall values our system has average performance. This is not bad because most systems that perform better than ours have participated in the OAEI from the beginning and they had more opportunity to tune and improve their systems over time. The comparative graph is depicted on the following Figure 5.

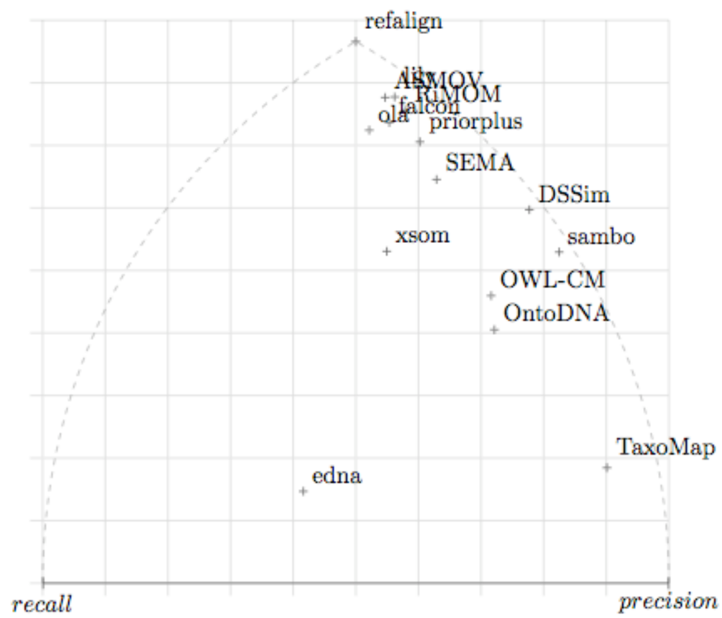

Figure 5

With the anatomy track our mapping performance was not satisfactory compared to other approaches. The main reason is the inefficient use of the background knowledge. The anatomy track is a medical ontology and the use of domain related background knowledge seems to be a crucial point in matching biomedical ontologies and the additional effort of exploiting this knowledge pays off.

With the directory track our precision result (see Figure 6) was among the bests however considering the recall we still have a lot of room for improvement.

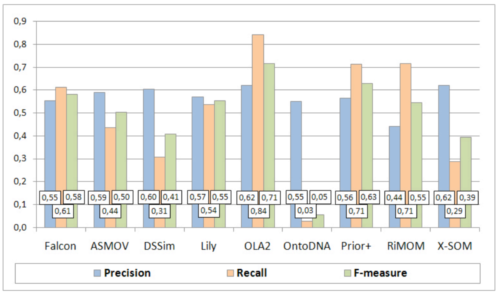

Figure 6

11 Applications of our Solution

Our proposed research objective goes beyond the ontology-mapping problem. Our aim is to provide a generic set of system components that can be used for Semantic data and application integration, which is an emerging solution for integrating multiple applications, databases and processes in the enterprise environment. As the Semantic Web initiatives can be seen as the next step in the evolution of data management, which provide the ability to share and analyze data stored by independent applications, semantic data and application integration can also be used in environment where it is necessary to discover relationships across different databases, business applications and Web services. One application area where we have already proposed our techniques is the Semantic Web Services [13] that is based on the loosely coupled architecture model namely the Service Oriented Architecture (SOA). This is an important area because it shows that the applicability of our solution cannot only be foreseen in the data application but the whole enterprise application domain (EAI).

12 Other Activities

12.1 Seminar

The second year seminar took place on Wednesday 17 January 2007.

12.2 Conferences/workshops

In ***2006*** I have attended 2 conferences and presented the following papers:

Nagy, M., Vargas-Vera, M., Motta, E., (2006) Uncertainty Handling in the Context of Ontology Mapping for Question Answering.. AAAI-2006 Symposium on Semantic Web for Collaborative Knowledge Acquisition.

Nagy, M., Vargas-Vera, M., Motta, E., (2006) DSSim-ontology Mapping with Uncertainty. International Worshop on Ontology Matching OM-2006, collocated with the 5th International Semantic Web Conference ISWC-2006.

The relevant EKAW workshop was cancelled but the following poster submission was accepted:
Nagy, M., Vargas-Vera, M., Motta, E., (2006) Similarity Mapping with Uncertainty for Knowledge Management of Heterogenous Scientific Databases in a Distributed Ontology-Mapping Framework. Poster, the 15th International Conference on Knowledge Engineering and Knowledge management (EKAW-2006)., Podebrady, Czech Republic.

In ***2007*** I have attended 2 conferences and presented the following papers:

Nagy, M., Vargas-Vera, M., Motta, E.,(2007), Multi-agent ontology mapping with uncertainty on the semantic web , In Proceedings of IEEE 3rd International Conference on Intelligent Computer Communication and Processing September 6-8, 2007 Cluj-Napoca, Romania

Nagy, M., Vargas-Vera, M., Motta, E.,(2007), Uncertain Reasoning for Creating Ontology Mapping on the Semantic Web, In proceedings of the workshop on Uncertainty Reasoning for the Semantic Web (URSW) in 6th International Semantic Web Conference (ISWC) Busan, Korea

Nagy, M., Vargas-Vera, M., Motta, E.,(2007), DSSim - managing uncertainty on the semantic web, The Second International Workshop on Ontology Matching collocated with the 6th International Semantic Web Conference ISWC-2007 November 11, 2007: BEXCO, Busan, Korea

Poster submission:

Nagy, M., Vargas-Vera, M., Motta, E.,(2007), Uncertain Reasoning in Multi-Agent Ontology Mapping on the Semantic Web, IEEE CS volume of the 6th Mexican International Conference on Artificial Intelligence, November 4-10, 2007 Aguascalientes, Mexico.

In ***2008***(till the end of July) I have attended 1 conference and presented the following papers:

Nagy, M., Vargas-Vera, M., and Motta, E. (2008) Feasible Uncertain Reasoning for Multi Agent Ontology Mapping, IADIS International Conference-Informatics 2008, Amsterdam, The Netherlands

13 References

[1] Heß A. (2006). An Iterative Algorithm for Ontology Mapping Capable of Using Training Data, Proceedings of the 3rd European Semantic Web Conference(ESWC 2006), Budva, Montenegro.

[2] Bonifacio M., Bouquet P., Traverso P. (2002). Enabling Distributed Knowledge Management: Managerial and Technological Implications . Technical Report DIT-02-069, Informatica e Telecomunicazioni, University of Trento, Italy.

[3] Smets, Ph. 1988. Belief functions, In Non-Standard Logics for Automated Reasoning: Academic Press, pp. 253–286.

[4] Zadeh, L. 1986. A simple view of the Dempster–Shafer theory of evidence and its implication for the rule of combination, AI Magazine 7, 85–90.

[5] Ramchurn, S. D., Huynh, T. D. and Jennings, N. R. 2004. The Knowledge Engineering Review, Volume 19: 1-25.

[6] Blok S., Medin D.L., Osherson D. (2003). Probability from similarity. In Proceedings of AAAI Conference on Commonsense reasoning, Stanford University, US.

[7] Ding C.H. (1999). A Similarity-based probability model for Latent Semantic Indexing. In Proceedings of SIGIR'99, volume 1, pages 58-65.

[8] Juslin P., Nilsson H., Olsson H. (2001). Where Do Probability Judgments Come From? Evidence for Similarity–Graded Probability, In Proceedings of 23rd Annual conference of the cognitive science society, Edinburgh, Scotland

[9] Kahneman D. et al. (1982). Judgment under uncertainty: Heuristics and Biases ,Conservatism in Human Information Processing,359, 361–69.

[10] Besana P.(2006). A Framework for Combining Ontology and Schema Matchers with Dempster-Shafer, Poster in the International Worshop on Ontology Matching OM-2006, collocated with the 5th International Semantic Web Conference ISWC-2006

[11] Smets P. (1990). The combination of evidence in the transferable belief model, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 12 , Issue 5: 447 – 458

[12] Yaghlane, B.B., Laamari, N. (2007) Owl-cm: Owl combining matcher based on belief functions theory. In Proceedings of the 2nd International Workshop on Ontology Matching

[13] Vargas-Vera M., et.al. (2008) Challenges on Semantic Web Services. Handbook of Research on Social Dimensions of Semantic Technologies and Web Services, http://handbooksemanticweb.ipca.pt/

14 Thesis outline

1. Introduction

1.1. Motivation

In this thesis we investigate the problem of ontology mapping using an agent based approach.

1.1.1. Semantic mapping

Semantic mapping generation algorithms between ontologies have inherent drawbacks when it comes to integrating data and information in real life scenario like question answering. Effectiveness can be improved but efficiency will worsen and vice versa. The reason for this is mapping process requires considerable background knowledge about the concepts, properties and its relation on the domain and different mapping algorithms use different information (e.g. predefined rules, machine learning) to assess similarity in advance instead of in real time. The problem is when mapping is made a priori, our real time question answering system will likely to fail when the resource (domain ontology, source) changes in the dynamic Web environment. Research questions that have been answered are:

- How to create mapping algorithms in a distributed environment that both effective and efficient and comparable to traditional solutions?
- How to replace the global knowledge with a distributed local knowledge of the multi agent system effectively?
- How trust in the different source information can be harnessed during the similarity combination process?

1.1.2. Uncertainty

Uncertainty handling requires human expert involvement in order to improve effectiveness of the system. Since uncertainty involves computing and combining probability distributions for all possible events any uncertainty handling formalisms are computationally expensive operations. Dempster-Shafer theory of evidence provides a promising alternative for reasoning under uncertainty. However the applicability in practical, real life scenarios is limited by the fact that combining the pieces of evidence with the Dempster’s rule of combinations suffers from the exponential growth of the state space therefore any system could be infeasible to build even with relatively small number of variables. Multi-variate Dempster-Shafer just worsen the situation since in this scenario each variable can have multiply values so the number of possible focal set is much bigger. Advanced optimisations and approximation helps to decrease the size of the state space and make the combination feasible although the applicability and feasibility of these methods and architectures have not been investigated in a distributed environment. Research questions that have been answered are:

- How Dempster-Shafer theory of evidence can be applied in a distributed multi agent environment for large complex domains?

- How mass probability can automatically be assessed from the similarity algorithms by specialized agents?
- How to apply traditional optimisation techniques in a distributed environment where pieces of the evidence are distributed between agents?

1.2. Outline

This section will provide an overview of the thesis based on the main contributions.

2. Background

2.1. Information integration and ontology mapping

This section provides an overview of what different research communities have investigated concerning the information integration problem. Numerous different approaches are presented in a way how different information sources can be integrated.

2.2. Uncertainty and Subjective probability methods

This section reviews the current research addressing the qualitative reasoning and decision-making problem under uncertainty concentrating especially to the Dempster-Shafer theory.

3. Similarity and ontology mapping

3.1. Semantic and syntactic similarity

In this section we discuss how similarity can be assessed for ontology mapping in the context of question answering.

3.2. Distributed knowledge management and the importance of local knowledge

We show how distributed local background knowledge resolves the limitations of the traditional knowledge management systems which are based on a single schema in the context of ontology mapping for question answering.

4. Uncertainty

4.1. Belief and trust for conflicting beliefs

In this section we introduce our method to create belief about the correctness of the mapping, show how to resolve conflict before combining belief through introducing trust between agents.

4.2. Optimisation of belief combination

In this chapter we will introduce our proposed optimisation methods that make belief combination possible in real time question answering system. Algorithms for variable elimination sequence in a distributed environment are presented and algorithms for distributed valuation network optimisations are discussed.

5. System

5.1. Architecture and Conceptual framework

This section describes the specific techniques that have been used and integrated into our system. Further it introduces the main components of the system

5.2. Interface

In this section I will discuss the level of interaction with the users that is necessary for the effective mapping composition.

5.3. Performance benchmarks

This section provides an overview of the performance benchmarks that have been carried out in order to measure the real time response time of our system.

5.4. Sample scenario

We provide a sample scenario, which helps the reader to better understand our proposed algorithm.

6. Evaluation

All the sections will describe the qualitative comparison of our system with other systems that were participated in the Ontology Alignment Initiative.

6.1. Evaluation using data set OAEI-2005
6.2. Evaluation using data set OAEI-2006
6.3. Evaluation using data set OAEI-2007
6.4. Evaluation using data set OAEI-2008
6.5. Evaluation using data set OAEI-2009
6.6. Overall conclusions from Evaluation

7. Conclusions

This chapter will contain all the conclusions that will be drawn from the research.

8. Discussion and summary of contribution

Reviews and summarizes the research contributions. Implications for related areas and possibilities for future work are discussed.

The foreseen contributions of my Ph.D. research can be grouped into two main areas:

1. Ontology mapping with multi agent for system integration in the context of Query answering:
    - Establishing an effective ontology concept similarity measures and combination algorithms that incorporates uncertainty and based on concept name, property hierarchy structure and instance values in a distributed environment.
    - Assessing and incorporating trustworthiness of the sources into the combination rule based on domain, author and time related information.
    - Agent communication strategies for combining pieces of evidence (Dempster-Shafer) where the evidence is distributed among specialized agents.
2. Effective reasoning under uncertainty in the context of ontology mapping using Dempster Shafer theory of evidence.
    - Algorithms for determining the variable elimination sequence when the joint tree is distributed between the agents.
    - Algorithms for building up a joint tree in the distributed environment.

15 Timetable for the third year

Planned activity	Time Frame
Result dissemination and networking (workshop, conferences)	October 2008 - December 2008
Comparison of algorithms for variable elimination in a distributed environment.	January 2009 - April 2009
OAEI 2009 participation + writing up PhD Thesis	May 2009 - September 2009
Prototype system development and integration of various components	October 2009 – January 2010
Finalise PhD Thesis	January 2010 – July 2010