Ontology-based question answering in a federation of university sites: the MOSES case study

Paolo Atzeni (3), Roberto Basili (1), Dorte H. Hansen (2), Paolo Missier (3),
Patrizia Paggio (2), Maria Teresa Pazienza (1), Fabio Massimo Zanzotto (1)

(1) Dip. di Informatica Sistemi e Produzione, University of Rome "Tor Vergata"
{basili,pazienza,zanzotto}@info.uniroma2.it
(2) Centre for Language Technology, University of Copenhagen
{patrizia,dorte}@cst.dk
(3) Dip di Informatica e Automazione, Universtità Roma Tre
atzeni@dia.uniroma3.it, pmissier@acm.org

Abstract

This paper deals with a new approach to ontology-based QA in which users ask questions in natural language to knowledge bases of facts extracted from a federation of Web sites and organised in topic map repositories. Our approach is being investigated in the context of EU project MOSES with the objective of developing an ontology-based methodology to search, create, maintain and adapt semantically structured Web contents according to the vision of the Semantic Web. MOSES is taking advantage of expertise coming from several fields: software agent technology, NLP, graph theory, text mining and data management. The test-bed chosen in the project is related to the development of an ontology-based knowledge management system and an ontology-based search engine that will both accept questions and produce answers in natural language for the Web sites of two European universities. After describing the requirements of the two user groups and proposing a classification of the questions the system is to support, the paper focuses on the interaction between NLP techniques and ontological knowledge, and on the use of ontology mapping for the treatment of federated questions.

Keywords

Question answering, natural language processing, ontology-based question analysis, ontology mapping, multi-linguality.

1 Introduction

Question Answering (QA) systems (as QA track of the Text Retrieval Conference (TREC-QA) competitions [21]), are able both to understand questions in natural language and to produce answers in the form of selected paragraphs extracted from very large collections of text. Generally, they are open-domain systems, and do not rely on specialised conceptual knowledge, while using a mixture of statistical techniques and shallow linguistic analysis. Ontological Question Answering systems, e.g. [23,24] propose to attack the problem by means of an internal unambiguous knowledge representation. As any knowledge intensive application, ontological QA systems have as intrinsic limitation related to the small scale of the underlying syntactic-semantic models of natural language.

While limitations are well-known, we are still questioning if any improvement has occurred since the development of the first ontological QA system LUNAR. Several important facts have emerged that could influence related research approaches:

- a growing availability of lexical knowledge bases that model and structure words: WordNet [17] and EuroWordNet [22] among others; some open-domain QA systems have proven the usefulness of these resources, e.g. WordNet in the system described in [12];

- the vision of a Web populated by "ontologically" tagged documents which the semantic Web initiative has promoted; in case this vision becomes a reality, it will require a world-wide collaborative work for building interrelated "conceptualisations" of domain specific knowledge;
- the trend in building shallow, modular, and robust natural language processing systems [1, 13, 2, 7] which is making them appealing in the context of ontological QA systems, both for text interpretation [2] and for database access [19].

Given this background, we are investigating a new approach to ontology-based QA in which users ask questions in natural language to knowledge bases of facts extracted from a federation of Web sites and organised in topic map repositories [11]. Our approach is being investigated in the context of EU project MOSES[1], with the explicit objective of developing an ontology-based methodology to search, create, maintain and adapt semantically structured Web contents according to the vision of the Semantic Web. MOSES is taking advantage of expertise coming from several fields: software agent technology, NLP, graph theory, text mining and data management. The test-bed chosen in the project is related to the development of an ontology-based knowledge management system and an ontology-based search engine that will both accept questions and produce answers in natural language for the Web sites of two European universities. The challenges of the project are:

- building an ontological QA system;
- developing a multilingual environment which implies the ability to handle different languages, and, importantly, several conceptualisations.

This paper deals with some of the aspects of the MOSES methodology. Section 2 describes the requirements of the two user groups and proposes a structural and a linguistic classification of the questions the system is to support. Section 3 and 4 discuss the way in which the project intends to comply with the semantic Web vision, and raise the question of ontology mapping in a multilingual environment. Section 5 describes how question analysis is performed, focusing in particular on the interplay between NLP techniques and ontological knowledge. It also proposes an approach to the analysis of questions addressed to a federation of sites (federated questions) that relies on ontology mapping.

2 Requirements and Questions Classification Framework

Before discussing our approach to ontology-based question answering, we introduce the scope of this research. In the context of the MOSES project, user groups from an Italian and a Danish University were first asked to identify a common testbed for benchmarking the QA system under development, assuming their respective University web sites as the target information repositories.

Specifically, they were asked to provide both a conceptualisation of their respective domains, expressed using different ontologies, and a collection of reference questions regarding the web site contents, that would serve both as requirements for the design of the QA system, and as testbed to benchmark its effectiveness.

Interestingly, despite the differences in information content (Electronic and Computer Engineering for the Italian web site, Humanities for the Danish site), the groups found a common ground of concepts that would not coincide, but overlap partially, making a mapping between the two ontologies possible. The issue of ontology mapping is discussed in Section 4.

As a consequence, the set of test questions were also comparable. Questions were chosen with representative users in mind (students, researchers, administrative staff -- eg "In which classroom are the Database Classes held this semester" is a common student question), according to three main requirements.

The first is a requirement on question content. In order to prove a clear added value to using a QA system as opposed to a traditional search engine, or plain site navigation, for each question the current "path to answer" was identified in order to derive a quantitative effectiveness benchmark for the system.

[1] MOSES is a cooperative project under the 5th Framework Programme. The project partners are FINSA Consulting, MONDECA, Centre for Language Technology, University of Copenhagen, University of Roma Tre, University of Roma Tor Vergata and ParaBotS.

The second requirement concerns question structure, namely questions should belong to well-defined classes with respect to both the *type* of information sought, and on the *type of linguistic analysis* required to interpret the question.

The third requirement states the need to support *federated questions*, that is, questions that span multiple content repositories (in our case, the two University test sites) in addition to local questions. For instance, asking for all active research on a given topic Europe-wide may well involve all University sites that may potentially provide relevant information. Issues raised by this requirement concern both the linguistic level and the architecture design. On one hand, questions posed in one language must be forwarded to sites that cannot interpret questions in that language, and answers provided in one language by a foreign site should be made readable to the local users. Thus, federated questions bring along the additional requirement of multi-linguality, On the other hand, an infrastructure for managing distributed QA must be in place. The former issue is briefly touched upon in Section 4, while the second is beyond the scope of this paper.

A fourth requirement on multi-lingual questions was of course implicit in the use of the groups' respective languages. All these requirements specifically push for both linguistic and ontological knowledge.

The collection of questions resulting from the users' effort, each described by metadata regarding both its structure and content, is proving effective in testing a number of key features derived from the requirements in the prototype system, including:

- building local ontologies by specialising and extending existing off-the-shelf ontologies, as discussed in Section 3.2;
- creating a mapping between different ontologies defining a similar semantic domain;
- providing ontology-based natural language question analysis with multi-lingual support.

The structural and linguistic classification framework for questions is key to making sure that the test bed covers all useful question types. Therefore, we devote the rest of this section to describing the framework in some detail.

2.2 Structural Classification of Questions

The structure of a question can be classified described according to various criteria that relate to specific issues in answering it.

The first criterion is the structure of the expected answer. The simplest format for a response is an atomic data item, like a person's name or phone number. This is the simplest response usually provided by a traditional database interface. Some simple type information, eg “email address”, “url” is often associated to the raw data items. More structured or semantically rich information includes:

- A Web *page*, or a document, or a document fragment. This is the standard output from search engines (with ranking);
- A *tuple* of data, resembling relational records, containing items that are atomic or themselves structured;
- A *concept*, eg a node in an ontology ("person" as opposed to an instance "Prof. John Doe"). Of course, only ontology-based systems are able to provide this level of information

Additionally, the cardinality of the returned item set matters:

- A list of items (for example, a list of course types), possibly ordered;
- A list of pages, possibly ranked by a search engine;
- A list of tuples, or concepts (essentially, a table, possibly nested): a list of the professors, each with the courses he/she teaches and a link to the home page.

A more general form of answer is a piece of natural language, which can be generated on the basis of a structured representation of concepts and relations, or a piece of text which is extracted directly from a Web page (in which case we have an overlap with one of the cases listed above).

The second criterion concerns locality, i.e., whether the question is federated, as defined earlier, or local to a single site.

The third criterion is on the actual structure of the source information, and it concerns how the information needed to answer the query is organised in the sites. We have three major cases:

- *Information directly provided by the site(s) as a whole:* this happens when the information is already on a page (or on a set of pages, if we are looking for a set of answers) and needs only to be extracted from there.
- *Information available in the sites and correlated but not aggregated:* in this case, the information of interest is spread over a set of pages, related by means of links to be followed to reach them.
- *Information available in the site in a unrelated way:* this case arises when the information is spread over pages that are implicitly but not explicitly correlated (for example, the correlation could be established by means of common values that appear in the pages but are not supported by links).

Finally, we classify questions according to their "current path to answer", of the following possible types:.

- *Explicit service:* this is the case for example if the site offers an address book or a search facility for finding professors.
- *One-way navigation:* here the user has to navigate, but without the need for going back and forth (no “trial and error”).
- *Navigation with backtracking:* this is the case when the navigation requires multiple attempts (for example, I am looking for a professor, but I do not know which department he/she belongs to, and there is no index).

2.3 Linguistic Classification of Questions

To facilitate recognition of what are the relevant expressions to be encoded in the linguistic interface, the questions have also been analysed and classified with a view to their syntactic form and their content. A classification often quoted is that in [15] which mainly builds on speech act theory. Another influential, more syntactically-oriented approach is that in [12] where to each syntactic category correspond one or several possible answer types, or focuses (a person, a date, a name, etc.).

From the point of view of the linguistic analysis, however, syntactic category and content are the central dimensions of sentence classification. Syntactic categories are e.g. *yes/no question, what-question, who-question, etc.* Subtypes relate to the position inside the question where the focus is expressed, e.g. depending on whether the wh-pronoun is a determiner, or the main verb is a copula. The content consists of concepts and relations from the ontology, the focus constraint[2] (the ontological type being questioned), and a count feature indicating the number of instances to be retrieved. Table 1 shows an example of linguistic classification. For each sentence type, several paraphrases are described.

[2] In the sense of [20].

FORM 1	
Input	Hvem underviser i filmhistorie (*Who teaches film history)*
Syntactic type	Who (Hvem)
Syntactic subtype	V ≠ copula
CONTENT	
Focus constraint	Teacher
Concepts	Faculty Course.Name: *history of film*
Relations	TeacherOf(Faculty, Course)
Answer count	List

Table 1: Example of question classification

3 An Ontology-based Approach to Question Answering

In our ontological QA system, both questions and domain knowledge are represented using the same ontology description language. It is foreseen to develop the QA system in two steps. First a prototypical implementation is planned to answer questions related to the current "state-of-affairs" of the site to which the question is posed. In a second step, given a "federation" of sites within the same domain, we will investigate whether and how an ontological approach could support QA across the sites. Answering a question can then be seen as a collaborative task between ontological nodes belonging to the same QA system. Since each node has its own version of the domain ontology, the task of passing a question from node to node may be reduced to a mapping task between (similar) conceptual representations. To make such an approach feasible, a number of difficult problems must still be solved. In this paper, we will provide details on how:

- to build on existing ontologies and interface between them and language resources;
- to interpret questions wrt the ontological language;
- to model the mapping task for federated questions.

3.2 Building on off-the-shelf semantic Web Ontologies

One of the results of the Semantic Web initiative will be the production of many interrelated domain-specific ontologies that provide the formal language for describing the content of Web documents. In spite of the freedom allowed in the production of new conceptualisations, it is reasonable to expect that a first knowledge representation jungle will leave room to a more orderly place where only the more appreciated conceptualisations have survived. This is a prerequisite for achieving interoperability among software agents. In view of this, and since publicly available non-toy ontology examples are already available, the effort of adapting an existing ontology to a specific application is both useful and possible.

Ontologies for the Semantic Web are written in formal languages (OWL, DAML+OIL, SHOE) that are generalisations/restrictions of Description Logics [4]. TBox assertions describe concepts and relations. A typical entry for a concept is:

ID	Course
Label	Course
Subclassof	Work

Table 2. A semantic concept

where ***ID*** is the concept unique identifier, ***label*** is the readable name of the concept, ***subclassof*** indicates the relation to another class. As the label has the only purpose of highlighting the concept to human readers, alternative linguistic expressions are not represented. Rather, this piece of information is recorded in a lexical data base like WordNet. The problem is even more obvious when considering relationships.

ID	teacherOf
Label	Teaches
Domain	#Faculty
Range	#Course

Table 3. A semantic relationship

In Table 3, *domain* and *range* contain the two concepts related to the described binary relation. The label *Teaches* does not mention alternative linguistic expressions like: *#Faculty* **gives** *#Course* or *#Faculty* **delivers** *#Course*, etc.

For the ontology producers, only one concept or relation name is sufficient. Synonymy is not a relevant phenomenon in ontological representations. In fact, it is considered a possible generator of unnecessary concept name clashes, i.e. concept name ambiguity. Conceptualisations (as in tables 2 and 3) are inherently weak whenever used to define linguistic models for NLP applications. Interpreting questions like:

(1) Who **gives**/**teaches** the database **class**/**course** this year?

with respect to a university domain ontology means in fact mapping all the questions onto the concepts and relations in Table 3. There is a gap to be filled between linguistic and ontological ways of expressing the domain knowledge.

3.3 Linguistic Interfaces to Ontologies

In developing an ontological QA system, the main problem is to build what we call the "linguistic interface" to the ontology which consists of all the linguistic expressions used to convey concepts and relationships. To make this attempt viable, we are currently studying methods to automatically relate lexical knowledge bases like WordNet [17] to domain ontologies [5] and to induce syntactic-semantic patterns for relationships [6].

The linguistic interface constitutes the basis on which to build the semantic model of the natural language processing sub-system. One way of conceiving such a model is in terms of syntactic-semantic mapping rules that apply to alternative expressions of the same conceptual knowledge. The amount of syntactic analysis such rules foresee will vary according to the approach chosen.

4 Ontology Mapping in a Multilingual Environment: challenges

The conceptualisation of the university world as it appears in the DAML+OIL ontology library is an interesting representation for the application scenarios targeted in MOSES (i.e. *People/Course/Research*). Described classes and relations cover in fact, at least at a high level, most of the relevant concepts of the analysed scenarios. Such an ontology has been adapted to develop conceptualisations for each of the two national university sub-systems (i.e. Italian and Danish) while providing additional information required for answering the input questions. This is temporal information or other kind of information at a borderline with the domain (e.g. concepts related to the job market). A first important matter we have dealt with is the language. Whereas concept and relation labels in the Italian ontology are expressed either in English (for concepts directly taken from the original source) or in Italian, in the Danish counterpart all labels are in Danish. This means that a mapping algorithm making use of string similarity measures applied to concept labels will have to work with translation, either directly between the two languages involved, or via a pivot language like English. The goal would be to establish correspondences such as 'Lektor' ↔ ('AssociateProfessor') ↔ 'ProfessoreAssociato'.

Another problem is related to structural differences: not all the nodes in an ontology are represented also in the other and vice-versa, moreover nodes that are somehow equivalent, may have different structural placements. This is the case for the 'Lektor'/'ProfessoreAssociato' pair just mentioned: in the Danish system, 'Lektor' is not a subclass of 'Professor', although "associate professor" is considered a correct translation.

5 Question Analysis

Question analysis is carried out in the MOSES linguistic module associated with each system node. To adhere to the semantic Web approach, MOSES poses no specific constraints on how the conceptual representation should be produced, nor on the format of the output of each linguistic module. The agent that passes this output to the content matcher (an ontology-based search engine) maps the linguistic representation onto a common MOSES interchange formalism (still in an early development phase). Two independent modules have been developed for Danish and Italian language analysis. They have a similar architecture (both use preprocessing, i.e. POS-tagging and lemmatising, prior to syntactic and semantic analyses), but specific parsers. Whereas the Danish parser, an adapted version of PET [9] produces typed feature structures [10], the Italian one outputs quasi-logical forms. Both representation types have proven adequate to express the desired conceptual content. As an example, the Italian analysis module is described below[3].

5.2 Analysis of Italian Questions

Analysis of Italian questions is carried out by using two different linguistic interpretation levels. The syntactic interpretation is built by a general purpose robust syntactic analyser, i.e. Chaos [7]. This will produce a Question Quasi-Logical Form (Q-QLF) of an input question based on the extended dependency graph formalism (XDG) introduced in [7]. In this formalism, the syntactic model of the sentence is represented via a planar graph where nodes represent constituents and arcs the relationships between them. Constituents produced are chunks, i.e. kernels of verb phrases (VPK), noun phrases (NPK), prepositional phrases (PPK) and adjectival phrases (ADJK). Relations among the constituents represent their grammatical functions: logical subjects (lsubj), logical objects (lobj), and prepositional modifiers. For example, the Q-QLF of the question

(2) Chi insegna il corso di Database?
(Who teaches the database course?)

is shown in Figure 1.

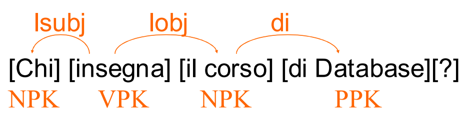

Figure 1 A Q-QLF within the XDG formalism

Then a robust semantic analyser, namely the Discourse Interpreter from LaSIE [14] is applied. An internal world model has been used to represent the way in which the relevant concepts (i.e. objects) and relationships (i.e. events) are associated with linguistic forms (see Figure 2). Under the object node, concepts from the domain concept hierarchy are mapped onto synsets (sets of synonyms) in the linguistic hierarchy EWN (i.e. the EuroWordNet.base concepts). This is to guarantee that linguistic reasoning analysis is made using general linguistic knowledge.

[3] The Danish system will be described in an extended version of the paper.

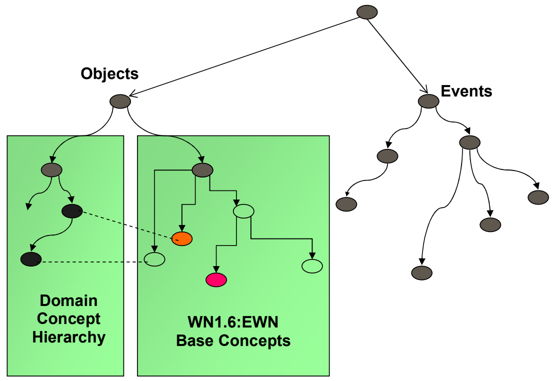

Figure 2 The world model taxonomy

The association of objects and events with linguistic forms is used in matching rules as shown in Figure 3. The rule expresses the fact that, if any word like *tenere*, *insegnare* or *fare* is encountered in relation with a *human_1* (represented by the base concept *ewn4123*) and the word *education_1* (*ewn567704*), the relation ***teacherOf*** can be induced.

```
TEACH_EVENT ==> teach_course.
        teach_course ==> tenere v insegnare v fare.

props(teach_course(E),[
        (consequence(E,
        [relation(E,teacherOf),r_arg1(E,X),r_arg2(E,Z)]     ):-
                nodeprop(E,lsubj(E,X)), X <- ewn4123(_),          /* human_1 */
                nodeprop(E,lobj(E,Z)), Z <- ewn567704(_)          /* education_1 */
        )
]).
```

Figure 3 Example of syntactic-semantic interpretation rule

The analysis resulting for sentence (2) is then:

```
focus(e2),
relation(e1,teacherOf),
r_arg1(e1, person_dch(e2)),
r_arg2(e1,course_dch(e3)),
relation(e4,hasSubject),
r_arg1(e4, course_dch(e3)),
r_arg2(e4,topic_dch("Database")).
```

This means that the user is interested in a person, the entity *e2* of the class person_dch, that is in a relation teacherOf with the entity *e4* (instance of the class course_dch), that is in turn related by hasSubject with the topic (i.e. topic_dch) "Database". This result can be passed on to the content matcher.

5.3 Handling Federated Questions

Now we want to extend this approach to question analysis in order to manage federated questions. A possible solution would be sending the natural language question to several nodes and let each node interpret it against its own domain knowledge. This is unfeasible in a multilingual environment. The solution we are investigating is based on the notion of ontology mapping. Let us consider the case of a student questioning not only the Danish but also the Italian site (by selecting specific modalities for entering questions):

(3) Hvem er lektor i fransk?
(Who is associate professor of French?)

As the question is in Danish, it has to be analysed by the Danish analysis component, which will produce a semantic interpretation roughly equivalent to the following term:

all(x) (lektor(x) & CourseOffer(x,y) & Course(y) & Name(y, French))[4]

Since all concepts and relations come from the Danish ontology, it is not a problem to query the Danish knowledge base for all relevant examples. In order to query the Italian knowledge base, however, equivalent concepts and relations must be substituted for those in the "Danish" interpretation. The corresponding Italian representation is:

all(x) (ProfessoreAssociato(x) & TeacherOf(x,y) & Course(y) &
Subject(y, French))

The first problem is establishing a correspondence between 'lektor' and 'ProfessoreAssociato', which as shown in the ontology fragments in Figure 4, are not structurally equivalent.

As suggested in [18,16], equivalence relations must be established by considering *is-a* structures and lexical concept labels together. In the example under discussion, an initial equivalence can be posited between the top nodes of the two ontology fragments, since they both refer explicitly to the original DAML+OIL ontology via a *sameAs* relation. However, none of the concept labels under 'Faculty' in the Italian ontology are acceptable translations of 'Lektor', nor do any of the nodes refer to common nodes in a common reference ontology. Thus, the matching algorithm must search further down for en equivalent concept by considering possible translations of concept labels and testing the relations that equivalence candidates participate in. Thus, distance from a common starting node, lexical equivalence and occurrence in similar relations are all constraints to be considered.

The same problem of finding a correct mapping applies to relations. In this case, we must be able to discover that CourseOffer and TeacherOf represent the same relation. For instance we can rely on the fact that they have both two roles, and the concepts filling these roles, Faculty and Course (or rather the Danish and Italian equivalent concepts) correspond. Discovering similarities between relations, however, may be a much more complex task than shown in this example. In general, it assumes the ability to map between concepts.

6 Conclusion

Our focus in this paper has been, in the context of ontology-based QA, to discuss how to interface between ontology and linguistic resources on the one hand, and ontology and natural language questions on the other while remaining within a unique framework. We have described how multilingual input questions are analysed and classified to define the scope of the QA system, and explained the way in which NLP techniques interact with the ontology to produce semantic representations of the questions to be fed to the system's content matcher. An interesting requirement to the framework we are developing is that it must deal with a multilingual environment. This in turn means supporting questions to federation of sites organised around local ontologies. It is shown in the paper that this issue can be addressed in terms of ontology mapping, and that mapping algorithms can benefit from looking not only at structural equivalence, but also at similarity of concept labels.

[4] All concepts and relations will in fact be expressed in Danish. Here, to facilitate non-Danish readers, we are using English equivalents with the exception of the concept 'Lektor' under discussion.

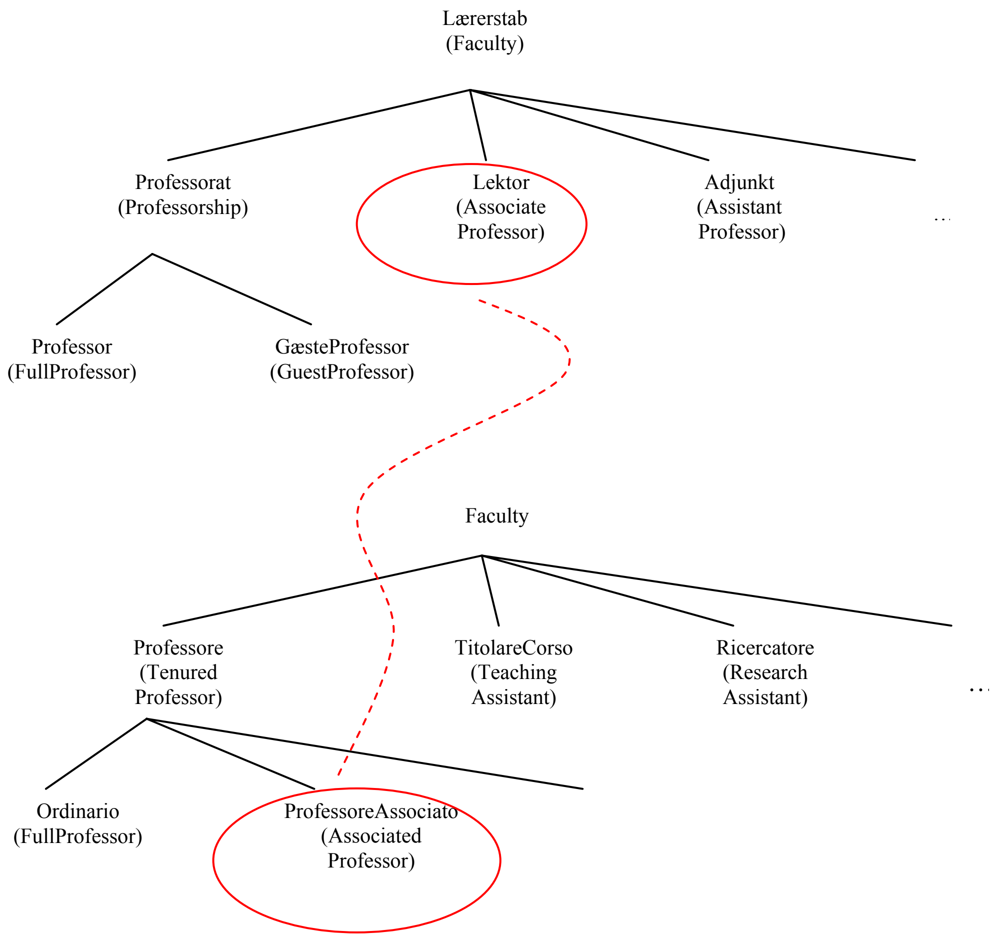

Figure 4: The “Faculty” Danish and Italian sub-ontologies

References

1. Steven Abney (1996) Part-of-speech tagging and partial parsing. In G.Bloothooft K.Church, S.Young, editor, Corpus-based methods in language and speech. Kluwer academic publishers, Dordrecht.
2. Salah Ait-Mokhtar and Jean-Pierre Chanod. (1997) *Incremental Finite-state parsing*. In Proceedings of ANLP97, Washington.
3. Andreasen, Troels, Per Anker Jensen, Jørgen F. Nilsson, Patrizia Paggio, Bolette Sandford Pedersen and Hanne Erdman Thomsen (2002) Ontological Extraction of Content for Text Querying, in *Natural Language Processing and Information Systems*, Revised Papers of NLDB 2002. Springer-Verlag, pp. 123–136.
4. Baader, F., D. Calvanese, D. McGuinness, D. Nardi, P.F. Patel-Schneinder, eds. (2003) *The Description Logics Handbook: Theory, Implementation, and Applications*, Cambridge University Press
5. Basili, Roberto, Michele Vindigni, Fabio Massimo Zanzotto (2003a) *Integrating ontological and linguistic knowledge for Conceptual Information Extraction*, Web Intelligence Conference, Halifax, Canada, September 2003
6. Basili, Roberto, Maria Teresa Pazienza, and Fabio Massimo Zanzotto (2003b) *Exploiting the feature vector model for learning linguistic representations of relational concepts* Workshop on Adaptive Text Extraction and Mining (ATEM 2003) held in conjuction with Europena Conference on Machine Learning (ECML 2003) Cavtat (Croatia), September 2003
7. Basili, Roberto and Fabio Massimo Zanzotto (2002) *Parsing Engineering and Empirical Robustness* Journal of Natural Language Engineering 8/2-3 June 2002

8. Burger, John *et al* (2002) *Issues, tasks and program structures to roadmap research in question & answering* (Q&A). NIST DUC Vision and Roadmap Documents, http://www-nlpir.nist.gov /projects/duc/roadmapping.html.
9. Callmeier, Ulrich (2000) PET – a platform for experimentation with efficient HPSG processing techniques. In Flickinger, D., Oepen, S., Tsujii, J. and Uszkoreit, H. (eds.) *Natural Language Engineering. Special Issue on Efficient Processing with HPSG*. Vol. 6, Part 1, March 2000, 99–107.
10. Copestake, Ann (2002) *Implementing Typed Feature Structure Grammars*. CSLI Publications. Stanford University.
11. Garshol, Lars Marius (2003) Living with Topic Maps and RDF. Technical report. http://www.ontopia.net/topicmaps/materials/tmrdf.html.
12. Harabagiu, Sanda, Dan Moldovan, Marius Paca, Rada Mihalcea, Mihai Surdeanu, Rhzvan Bunescu, Roxana Girju, Vasile Rus, and Paul Morrescu (2001) *The role of lexico-semantic feedback in open-domain textual question-answering*. In Proceedings of the Association for Computational Linguistics, July 2001.
13. Hobbs, Jerry R., Douglas E. Appelt, John Bear, David Israel, Megumi Kameyama, Mark Stickel, and Mabry Tyson (1996). *FASTUS: A cascaded finite-state transducer for extracting information from natural-language text*. In Finite State Devices for Natural Language Processing. MIT Press, Cambridge, MA.
14. Humphreys, K., R. Gaizauskas, S. Azzam, C. Huyck, B. Mitchell, H. Cunningham, and Y. Wilks (1998) *University of sheffield: Description of the LASIE-II system as used for MUC-7*. In Proceedings of the Seventh Message Understanding Conferences (MUC-7). Morgan Kaufman, 1998.
15. Lauer, Thomas W., Eileen Peacock and Arthur C. Graesser (eds.) (1992) *Questions and Information Systems*. Hillsdale, NJ: Lawrence Erlbaum.
16. Meadche, Alexander and Steffen Staab (2001) *Comparing Ontologies-Similarity Measures and Comparison Study*, Internal Report No. 408, Institute AIFB, University of Karlsruhe, Germany, 2001
17. Miller, George A. (1995) WordNet: A lexical database for English. Communications of the ACM, 38(11):39--41, 1995.
18. Pazienza, Maria Teresa and Michele Vindigni (2003) *Agent-based Ontological Mediation in IE systems* in M.T. Pazienza ed. Information Extraction in the Web Era, LNAI 2700, Sprinter Berlin 2003.
19. Popescu, Ana-Maria, Oren Etzioni and Henry Kautz (2003) Towards a Theory of Natural Language Interfaces to Databases, in *Proceedings of the International Conference on Intelligent User Interfaces* IUI 2003, pp. 149–157.
20. Rooth, M. (1992) A Theory of Focus Interpretation. In *Natural Language Semantics*, Vol. 1, No. 1, pp. 75-116.
21. Voorhees, Ellen M. (2001) The TREC question answering track. *Natural Language Engineering* 7(4), pp. 361–378.
22. Vossen, Piek (1998) *EuroWordNet: A Multilingual Database with Lexical Semantic Networks* Kluwer Academic Publishers, Dordrecht, October 1998
23. Woods, W., R. Kaplan, and B. Nash-Weber (1972) *The Lunar Sciences Natural Language Information System: Final Report*. Technical Report, Bolt Beranek and Newman, Number 2378, June 1972.
24. Zajac, Remi (2001) *Towards Ontological Question Answering*, ACL-2001 Workshop on Open-Domain Question Answering, Toulose, France, 2001.