Learning Medical Ontologies from the Web

David Sánchez and Antonio Moreno

Intelligent Technologies for Advanced Knowledge Acquisition (ITAKA) Research Group
Department of Computer Science and Mathematics
Universitat Rovira i Virgili (URV)
Avda. Països Catalans, 26. 43007 Tarragona (Spain)
{david.sanchez,antonio.moreno}@urv.net

Abstract. The development of intelligent healthcare support systems always requires a formalization of medical knowledge. Domain ontologies are especially suitable for this purpose but their construction is, in most cases, manually addressed. This results in long and tedious development processes that hamper their real applicability. This is why there is a need of ontology learning methods that aid the ontology construction process. Considering the enormous amount of digital medical knowledge available freely on the Web, one may consider it as a valid source for developing knowledge acquisition systems. In this paper we offer an overview of an automatic and unsupervised method for learning domain ontologies from the Web. We also introduce its application over a specific medical domain in the frame of the K4Care European project.

Keywords: Ontology learning, web mining, knowledge acquisition, medical knowledge modelling.

1 Introduction

Medical ontologies are developed to solve problems such as the demand for reusing and sharing patient data or the transmission of these data [13]. The unambiguous communication of complex and detailed medical concepts is a crucial feature in current medical information systems. In these systems, several agents must interact in order to share their results and, thus, they must use a medical terminology with a clear and non-confusing meaning [9].

The development of these ontologies is a complex task: on the one hand, they are general enough to be required for achieving consensus between a wide community of users and, on the other hand, they are concrete enough to present an enormous diversity with hundreds of possible concepts to model.

Medical ontology engineering is typically addressed manually, requiring the intervention of medical specialists (which provide the medical knowledge) and knowledge engineers (which are able to formalize that knowledge). The necessary consensus is typically hampered by the difficulty of translating the shared world model of a medical community to the formal and explicit knowledge representation that an ontology definition requires. This produces long and tedious development stages that delay the applicability of the resulting ontologies.

D. Riaño (Ed.): K4CARE 2007, LNAI 4924, pp. 32–45, 2008.

Due to all these reasons, nowadays, there is a need of methods that can perform, or at least ease, the construction of medical ontologies. In this sense, *Ontology learning* is defined as the set of methods and techniques used for building from scratch, enriching, or adapting an existing ontology in a semi-automatic fashion using distributed and heterogeneous knowledge and information sources [9]. These methods allow a reduction in the time and effort needed in the ontology development process.

This data-driven knowledge acquisition process typically uses scientific texts, electronic dictionaries or medical repositories (such as UMLS). Considering the nature of those learning corpus (reduced scope, noise-free, trusted, structured), classical ontology learning methods have been designed [9].

In the last years, the growth of the medical information available on the Web provides users with a way for fast data access and information exchange. It is an invaluable tool for researchers, information engineers and health care companies and practitioners [8] for retrieving knowledge. These characteristics have motivated researchers [21] to consider the Web as a valid repository for *Information Retrieval* and *Knowledge Acquisition*. However, the extraction of information from web resources is a difficult task, due to their lack of semantic structure, noise, commercial bias and untrustworthiness, in addition to the ambiguity inherent to all resources written in natural language.

Despite all these shortcomings, the Web also presents characteristics that can be interesting for knowledge acquisition. As the number of resources available is so vast and the amount of people generating web pages is so enormous, it has been argued that the Web information distribution approximates the real distribution as used in society [5]. From the learning point of view, this is a very interesting characteristic and our motivation for using the Web as the source for knowledge acquisition.

So, in this paper, we present an overview of a novel approach for automatic domain ontology learning from the Web. Thanks to the amount of medical information available on the Web and the structured nature of medical knowledge, our method is especially suitable for learning medical ontologies. As a result of the application of this methodology over a medical domain, we introduce a case of study framed in the scope of the K4Care European research project. At the end, the main aim of this paper is to shown the usefulness of the developed automatic learning method to aid medical researchers in modelling knowledge.

The rest of the paper is organized as follows. Section 2 presents an overview of the main steps involved in the ontology construction process, introducing the learning techniques employed for knowledge acquisition. Section 3 gives a general vision of our approach for learning domain ontologies from the Web, including the incremental acquisition of taxonomic and non-taxonomic relationships and named entities. Section 4 presents and evaluates an example of the obtained results for a medical domain framed in the context of the K4Care European research project. The final section presents the conclusions and proposes lines of future work.

2 Ontology Learning Overview

In this section we introduce the ontology learning life-cycle, describing the main steps and ontological entities that should be considered during the ontology construction process. For each of them, the main learning techniques and hypothesis employed during the definition of our automatic methodology are introduced.

Ontologies are composed at least by *classes* (concepts of the domain), *relations* (different types of binary associations between concepts or data values) and *instances* (real world individuals). Formally, an ontology often boils down to an object model represented by a set of concepts or classes C, which are *taxonomically* related by the transitive *IS-A* relation H ϵ C x C and *non-taxonomically* related by named object relations R* ϵ C x C x String. On the basis of the object model, a set of logical axioms, A, enforce semantic constraints.

From the *Ontology engineering* point of view, there are several methodologies for constructing ontologies from scratch. In [9], an overview of the methods is presented. Analyzing them, the main steps and knowledge acquisition techniques employed for building ontologies are (see Fig.1):

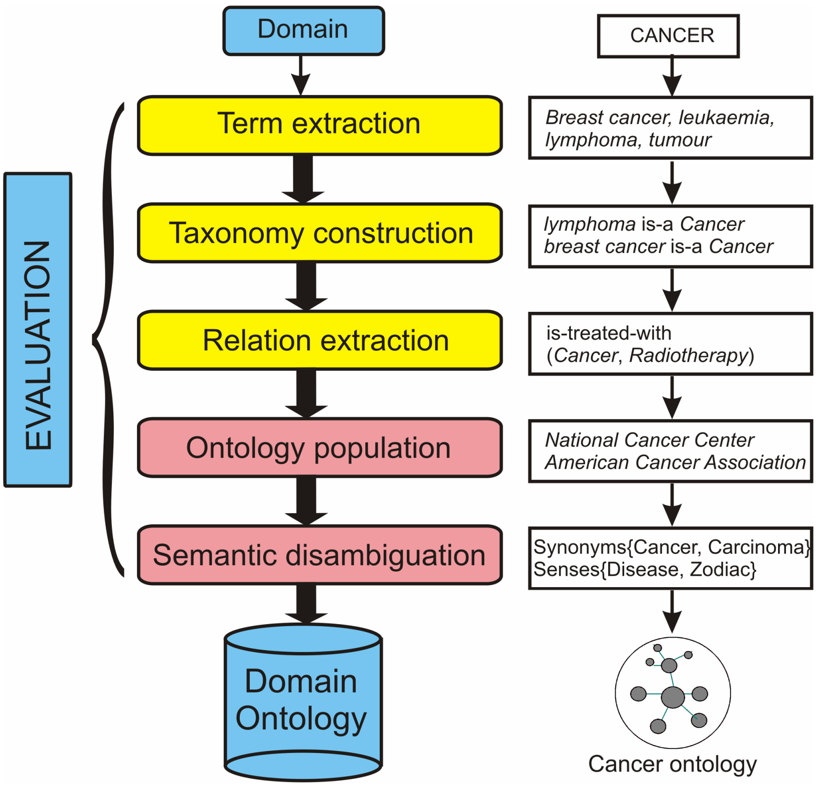

Fig. 1. General steps of the domain ontology learning process

- Extraction of terms that represent domain concepts, building a lexicon. Unsupervised approaches typically rely on statistical analyses about term co-occurrence [24]. They try to infer concept semantics by studying domain information distribution computed from a general corpus. The problems of computing robust measures and avoiding data sparseness are commonly addressed by using the Web. Concretely, highly valuable statistics can be obtained in an immediate way from the hit counts of web search engines if the appropriate queries are performed [27]. Thanks to the size and heterogeneity of the Web, those values are very robust, as they can approximate the true societal words usage [5].

- Construction of an initial taxonomy of concepts using *is-a* relations. From an unsupervised point of view, as stated in [6], three different learning paradigms can be exploited. First, some approaches rely on the document-based notion of term subsumption [25]. Secondly, some researchers claim that words or terms are semantically similar to the extent to which they share similar syntactic contexts [4]. Finally, several researchers have attempted to find taxonomic relations expressed in texts by matching certain patterns associated to the language in which documents are presented [1]. We have opted for this last approach because hyponymy detection linguistic patterns such as Hearst's [11] or Grefenstette's ones [10] can be used to construct Web Information Retrieval queries.
- Learning non-taxonomic relations. It is considered as the least tackled problem within ontology learning [14]. It appears to be the most intrincate task as, in general, it is less known how many and what type of conceptual relationships should be modelled. We have addressed the problem by extensively using verb phrases as the central point of a relation. From the ontology engineering point of view, verbs express a relation between two classes that specify the domain and range of some action or event [26]. Following the same philosophy as in the taxonomic case, we consider specific verb phrases as particular domain-dependent semantic patterns that express a particular non-taxonomic relationship [22]. Lightweight analytic procedures [19] and statistics compiled from querying a web search engine [27] complete the proposed non-taxonomic learning method.
- Ontology population by the detection of instances for the discovered concepts. We have limited this stage to the discovery of named entities. Similarly to the previous steps, we use language-dependent rules (capitalization for the English language) to detect proper names.
- Optionally, we can also treat semantic ambiguity in order to improve the quality of the results. We have developed complementary mechanisms to deal with polysemy and synonymy [23].
- Evaluation of the obtained results (concepts, instances and relationships). As ontological knowledge is non-uniquely expressible, the comparative evaluation of different approaches is difficult. For that reason, ontology learning evaluation is recognized to be an open problem [9]. In our case, as the quality of the final result will depend on the performance of every step of the learning process, specific evaluation methods for each one of them have been designed. Whenever a domain standard is available (e.g. MESH for the taxonomic case), results have been carefully compared. In other cases (as for the non-taxonomic relationships), an expert's opinion may be required.

3 Ontology Learning Methodology

In this section we offer an overview of the developed ontology learning method. Note that this section only represents an overview of the learning process, as our main objective is to introduce the usefulness of the developed methodologies in modelling domain knowledge. More details are offered in [22], [23] and [24].

The core of our Web-based approach covers the acquisition of domain terms and the definition of taxonomic and non-taxonomic relations. Its main advantage is the automatic and unsupervised operation, creating domain ontologies from scratch.

Even though we have developed individual methodologies for dealing with each learning step, they have been designed to be executed in an integrated and iterative way. Thus, each step can be bootstrapped with the knowledge acquired up to that moment. In this manner, new concepts and relationships can be used as seeds for further analysis. Through several iterations, the system incrementally constructs the semantic network of concepts composing the domain ontology.

As shown in Fig. 2, the learning process is divided in several phases.

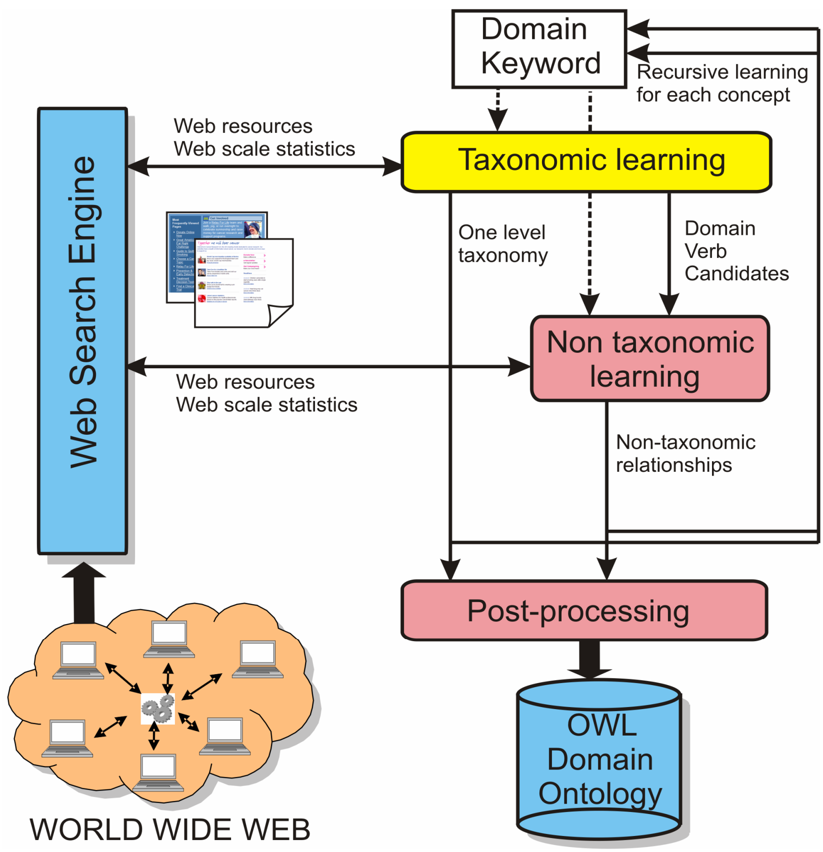

Fig. 2. Ontology learning methodology

The *Taxonomic learning* [24] starts from a user specified keyword (e.g. *cancer*) that indicates the domain for which the ontology should be constructed. The system starts by querying a web search engine to obtain a corpus of web documents to analyse. At this initial stage, only general queries using several domain-independent patterns for hyponymy detection (e.g. "*cancers such as*") are constructed. Web content is parsed in order to find matchings for those patterns and extract taxonomic candidates (e.g. *"cancers such as leukaemia or breast cancer"*).

Domain verbs found in the same context –sentence- as the pattern are also retrieved at this stage (e.g. *"cancer is associated with"*). Several iterations using different hyponymy detection patterns are performed in order to minimize language ambiguity and maximize the recall, and a final set of candidates is compiled.

An overview of the taxonomic learning process with an illustrative example is presented in Table 1.

Table 1. Heart's based learning overview: query, sample URL, sample web text (matching pattern in yellow), analysed sentences (valid candidates in yellow, candidate verbs in green), statistical analysis of candidates (selected ones in green)

Web Query	"cancers such as"
URL	http://www.dh.sa.gov.au/pehs/cancer-maps/cancer-maps-91-00.htm
Sample text	[...] There are several clear patterns which emerge on some of the maps. Firstly, ***cancers such as breast, melanoma and prostate cancer***, which require screening or a medical check for detection, almost always have higher incidence rates in high socio-economic status areas such as eastern and inner southern Adelaide. [...]
Analysed sentences	[ADVP Firstly/RB] ,/, [NP cancers/NNS] [PP such/JJ as/IN] [NP ***breast/NN*** ,/, ***melanoma/NN*** and/CC ***prostate/NN cancer/NN***] ,/, [NP which/WDT] *[VP require/VBP]* [NP screening/NN] or/CC [NP a/DT medical/JJ check/NN] [PP for/IN] [NP detection/NN] ,/, [ADVP almost/RB always/RB] ***[VP have/VBP]*** [NP higher/JJR incidence/NN rates/NNS] [PP in/IN] [NP high/JJ socio-economic/JJ status/NN areas/NNS] [PP such/JJ as/IN] [NP eastern/JJ and/CC inner/JJ southern/JJ Adelaide/NNP] ./.
Candidate evaluation (thres=1E-5)	Hits("cancers such as breast") = 12.774 Hits("breast") = 137.310.395 Score = 9.3E-5
	Hits("cancers such as melanoma") = 2.432 Hits("melanoma") = 864.002 Score = 2.4E-3
	Hits("cancers such as prostate cancer") = 1.827 Hits("prostate cancer") = 2.405.772 Score = 7.59E-4

Each taxonomic candidate is then evaluated using web-based statistical scores about term co-occurrence. New queries for web search engines are constructed in order to infer the degree of relatedness of a taxonomic candidate (e.g. *"breast cancer", "leukaemia"*) and the domain (e.g. *"cancer"*).

Web search hit counts are used to compute statistical scores (1). Those candidates with the higher scores are selected as valid taxonomic specialisations for the domain.

$$Score(Concept, domainKeyword) = \frac{hits("domainKeyword"\ AND\ "Concept")}{hits("Concept")} \quad (1)$$

In parallel, a procedure that detects named entities using capitalization heuristics is executed. It allows filtering the retrieved candidates by including real world individuals (e.g. *"American Cancer Association"*) as instances –and not incorrect subclasses- of the ontology.

At the end of all this process there is a one-level taxonomy with general terms (e.g. several types of *cancer*) and a set of verbs that have appeared in the same context – sentence- as the searched domain keyword (e.g. *is associated with, causes, is treated with*, etc.).

The next stage is the *Non-taxonomic learning* [22]. This process begins with the verb list compiled in the previous step, which is used as the knowledge base for the non-taxonomic learning. Each verb can be used as a bootstrap by constructing domain-related patterns (e.g. *"cancer is treated with"*) that are queried into the Web search engine. Additional web resources are retrieved and analysed to find verb-based pattern matchings (e.g. *"cancer" "is treated with" "radiotherapy"*). In order to minimize natural language ambiguity, only those sentences containing the pattern's instance that match with a set of simplicity rules are evaluated. Concretely sentences must be of the form:

<Sentence> [NP Subject] [VP Verb] ([PP Preposition]) [NP Object] </Sentence>

Similarly to the taxonomic case, candidates for non-taxonomic relations (e.g. *"radiotherapy"*) are ranked and selected using web-scale statistical scores (1). Finally, the verb phrase is used to link each pair of concepts, defining a set of domain binary relations.

An overview of the described process with an illustrative example is presented in Table 2.

Table 2. Non-taxonomic learning overview: query, sample URL, sample web text (matching sentence in yellow), analysed sentences (valid concept in yellow), statistical analysis of candidates (selected ones in green)

Web Query	"is associated with cancer"
URL	www.us.novartisoncology.com/info/understanding/preventing.jsp
Sample text	[…]***A high-fat diet*** *is associated with cancer* of the breast, uterus, and prostate. The guilty foods are eggs, fatty meats, high-fat salad dressings and cooking oils, and dairy products such as whole milk, butter, and most cheeses.[…]
Analysed sentences	[NP A/DT ***high-fat/JJ diet/NN***] [VP is/VBZ associated/VBN] [PP with/IN] [NP cancer/NN] [PP of/IN] [NP the/DT breast/NN] ,/, [NP uterus/NN] ,/, and/CC [NP prostate/NN] ./.
Candidate evaluation (thres=0.01)	Hits("high-fat diet") = 483.000 Hits("cancer" AND "high-fat diet") = 279.000 Score = 0.58

The two previous learning stages are *recursively* executed for each obtained concept (taxonomically –e.g. *"breast cancer"*- or non-taxonomically –e.g. *"radiotherapy"*- related). Each one becomes an individual seed for further analysis. Those new learning iterations can use the already acquired knowledge as a bootstrap to contextualize web queries and to obtain more concrete web resources.

The specific number of learning iterations, the amount of resources analysed at each step and the finalization of the recursive analysis is controlled by the algorithm itself. The system continuously monitors the learning performance by computing, at the end of each individual learning pass (i.e. the query and processing of a specific taxonomic or non-taxonomic pattern), the percentage of selected and rejected candidates according to the statistical scores. This value measures the learning throughput of a specific concept and pattern, allowing the system to self-control the learning process. On the one hand, the most productive ones –higher learning rage- are further evaluated by retrieving and analysing additional web resources. On the other hand, for the less productive ones –lower learning rate-, the process is finished and the next pattern and/or concept is taken.

Considering the higher degree of contextualization allowed by the bootstrapped knowledge (i.e. more concepts in web retrieval queries), the learning process is able to finish adequately as very little or no more resources or candidates can be retrieved/extracted for very specific queries.

At the end of this incremental learning process, we obtain a multi-level taxonomy in which each concept can be non-taxonomically related to other ones. An illustrative example of the kind of the structure that we are able to obtain is presented in the next section.

As a final step, a p*ost-processing* stage is introduced in order to detect implicit relationships (such as multiple inheritances), equivalencies, avoid redundancies and discover general domain features (concept attributes). In this manner we are able to obtain a more compact and coherent structure that becomes the final domain ontology.

4 Case of Study

In this section we introduce an example of application and the corresponding evaluation of results obtained by our learning method for a particular medical domain framed in the scope of the European project K4Care.

K4Care (http://www.k4care.net) aims to create, implement, and validate a knowledge-based healthcare model for the professional assistance to senior patients at home. This new Healthcare Model for home care will contribute to achieve a European standard supported by ICT technologies that improves the efficiency of the care services for all the citizens in the enlarged Europe. As shown in Fig. 3, K4Care relies on the definition of domain ontologies, Electronic Health-Care Records and Formal Intervention Plans.

In more detail, a specific Patient-Case Profile Ontology (CPO) is being constructed. It aims to structure the knowledge available about the care of patients. It combines diseases, syndromes, signs and symptoms, social issues, assessment tests, and interventions in order to define a knowledge model of how to deal with Home-Care Patients.

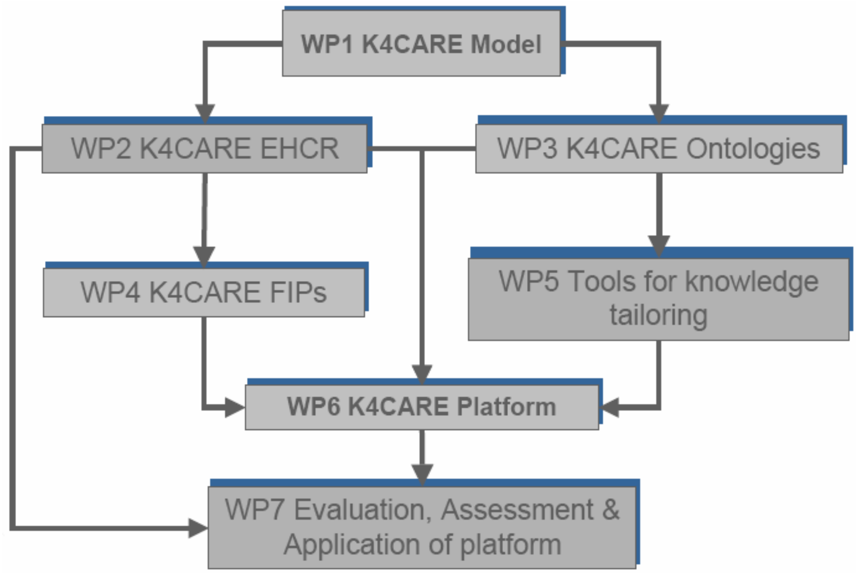

Fig. 3. K4Care work plan and dependencies. Ontologies are a fundamental part of the K4Care knowledge model.

More concretely, as described in [12], the information available in a patient's *Electronic Health-Care Record* (EHCR), combined with the results obtained for some clinical tests (*Comprehensive Assessment*), can be processed using the CPO as the knowledge base in order to infer the patient's syndromes or diseases. As a con- sequence, associated *Formal Intervention Plans* for the discovered pathologies can be used to aid (e.g. to suggest treatments, prescriptions, new medical tests, etc.) the healt- hcare providers in specifying the patient's particular treatment (*Individual Intervention Plans*).

The CPO is being currently defined manually from scratch, from the interaction of medical experts and knowledge engineers, supposing a considerable effort. Up to this moment, the ontology models the main entities that are relevant within the project scope.

During the earlier stages of the development, the ontology was heavily focused on the taxonomic aspect of the knowledge modelling (e.g. classification of different types of diseases), and offered a very little degree of general –non-taxonomic- semantic interlinkage between concepts (e.g. the symptoms corresponding to a disease), due to the inherent difficulty of manually modelling this kind of relationships.

In order to demonstrate the usefulness of an automatic ontology learning method in aiding the ontology development process, we have applied our method over a specific subdomain of the CPO. We have two main objectives. On the one hand, we aim to demonstrate the validity of our results by comparing them with the already modelled entities (mainly taxonomic). On the other hand, we argue how the manually composed ontology can be easily extended with the additional knowledge (mainly non-taxonomic) automatically acquired by our system.

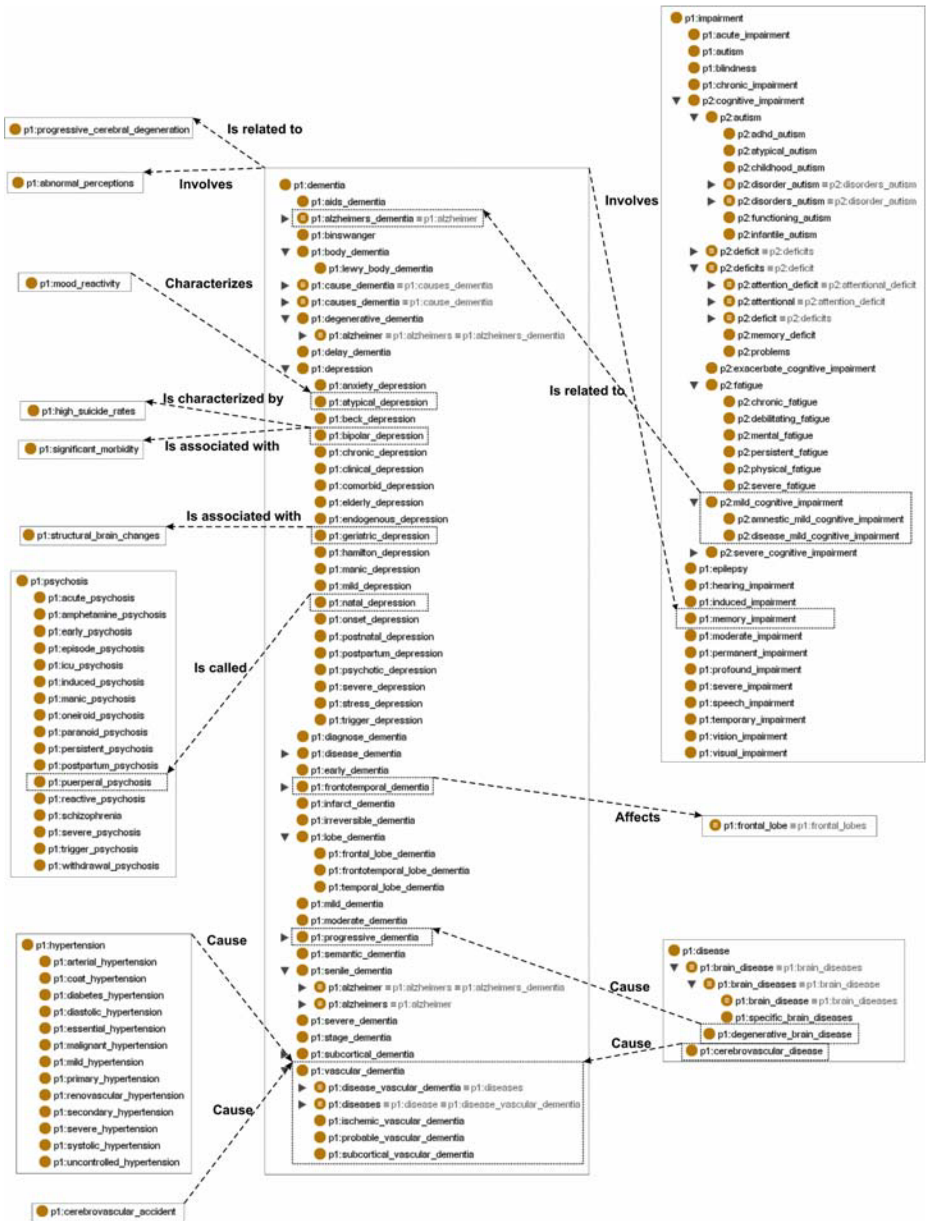

Fig. 4. Part of the domain ontology learned for the *Dementia* domain. A total of 236 classes, 57 instances, 99 non-taxonomic relations were obtained after 8 hours by analysing 21004 web sites.

In more detail, among the different entities modelled in the CPO, there are several diseases which are considered within the K4Care scope (typical pathologies of senior patients). The most exhaustively covered one is *Dementia*, for which several specialisations have been defined.

We executed our learning methodology for that domain. As a result, we obtained a Dementia ontology covering related classes, instances and taxonomic and non-taxonomic relationships. Most of the taxonomic relationships and some of the more relevant non-taxonomic ones are presented in Fig. 4. Considering the amount of discovered ontological entities, one can imagine the degree of human effort required to compile and structure them appropriately.

In order to evaluate the quality of these results in terms of *precision*, we compared them against a widely used medical standard repository (MESH http://www.nlm.nih.gov/mesh/MBrowser.html). We have queried the MESH browser to check if a discovered concept is present or not, obtaining a precision of 74% for the taxonomic case. Non-taxonomic relationships cannot be so easily checked as they are typically not modelled in standard classifications. A manual evaluation of the 99 discovered relationships measured a precision of 71.1%. In both cases, precision is high enough to consider the results as a reliable knowledge base for the domain.

Next, we compared the obtained ontology in terms of *recall* against the K4Care hand-made ontological specification. Considering that mainly taxonomic relationships are modelled in the CPO, including 15 types of diseases and 7 classes of dementia, we were able to retrieve 57% of them. The non-discovered ones are referred to the vaguest subclasses (e.g. *Mixed Type, Other Degenerative Dementia* and *Unspecified Dementia*) which are hardly distinguished from general adjectives. However, in total, we automatically discovered 25 direct subtypes of dementia, more than 200 related classes and 99 non-taxonomic relationships. Those last ontological entities are especially interesting due to the inherent difficulty of modelling non-taxonomic knowledge.

All those results were obtained in a completely automatic and unsupervised way, without requiring any kind of previous knowledge, search tuning according the domain or user's intervention. The system extensively queried a standard web search engine and analysed a large amount of web resources (21004). In any case, before its application in a real world environment, the ontology should be checked and filtered by a medical expert.

5 Conclusions and Future Work

Many knowledge acquisition approaches have been developed in the past. Different methodologies have been designed according to the knowledge source [18]: texts, dictionaries, knowledge bases, semi-structured data, relation schemas, etc. Considering the nature of those classical repositories, the common characteristics of classical knowledge acquisition methods are:

- Many of them [6, 16] use as learning corpus a reduced and pre-selected set of relevant documents for the covered domain. This approach solves some problems about untrustiness, lack of structure, noise and size that arise when developing an unsupervised, domain-independent Web-based approach.
- Most of the knowledge acquisition methodologies [1, 15] use predefined knowledge to some degree, like training examples, previous ontologies or semantic repositories. This fact hampers the development of domain-independent solutions, weakening the scalability and versatility of those systems in wide and heterogeneous environments like the Web.

- Most of them only cover the taxonomic aspect of the ontology learning process [14]. There have been very few attempts of non-taxonomic learning and, in many situations [17, 20], extracted relationships remain unlabelled.

On the contrary, we aim to obtain domain ontologies from scratch without any previous knowledge, adapting several classical techniques for knowledge acquisition (linguistic patterns, statistical analysis, etc.) to the special casuistry of the Web. We also cover all the main steps of the ontology learning process, configuring an integrated and intelligent learning approach.

Our approach is fully unsupervised. This is especially important due to the amount of available web resources, avoiding the need of a domain expert. The incremental learning method allows a dynamic adaptation of the evaluated corpus as new knowledge is acquired (bootstrap). Moreover, it has continuous feedback about the productivity of the learning task performed at each moment, guiding the learning to the most productive entities. In addition, the learning is automatic, allowing to easily perform executions at any time in order to retrieve updated results. This characteristic fits very well with the dynamic nature of the Web.

Domain ontologies are crucial in many knowledge intensive areas requiring interoperability such as the Semantic Web [3], e-commerce and e-medicine. From the presented results and posterior analysis, we can conclude that the use of automated ontology learning tools that are able to obtain with quite good accuracy (precision) a domain ontology in a few hours, can suppose a great help for ontology modellers. For the introduced example, the labour of specifying taxonomic entities can be reduced by more than a half. In addition, new ontological entities not yet considered (like new taxonomic terms and additional non-taxonomic relationships) are proposed.

Thanks to those advantages, ontology construction can be reduced from the fully manual ontology engineering effort -requiring an active participation of knowledge engineers- to a semi-automatic process which only requires refining a quite complete ontological structure. In this last case, ontologies can be evaluated and edited by the domain expert without advanced knowledge modelling skills.

As future research, we plan to apply our results to aid in the construction of the K4Care knowledge model. Other interesting syndromes, symptoms or diseases framed in the scope of the project can be further analysed. We would like to receive feedback for our results from expert medical partners of the K4Care project. This may give us an idea of the potential benefits and improvements that our solution may offer, such as the reduction in development time of the required knowledge structures.

Acknowledgements

The work has been supported by *Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya i del Fons Social Europeu* of Catalonia. Authors would also like to acknowledge the support of the K4Care European research project (IST-2004-026968) and the HYGIA project (TIN2006-15453-C04-01).

References

1. Ahmad, K., Tariq, M., Vrusias, B., Handy, C.: Corpus-based thesaurus construction for image retrieval in specialist domains. In: Sebastiani, F. (ed.) ECIR 2003. LNCS, vol. 2633, pp. 502–510. Springer, Heidelberg (2003)
2. Alfonseca, E., Manandhar, S.: An unsupervised method for general named entity recognition and automated concept discovery. In: Proceedings of the 1st International Conference on General WordNet (2002)
3. Berners-lee, T., Hendler, J., Lassila, O.: The semantic web. In: Scientific American (2001)
4. Bisson, G., Nedellec, C., Cañamero, D.: Designing Clustering Methods for Ontology Building. The Mo'K Workbench. In: Proceedings of the Workshop on Ontology Learning, 14th European Conference on Artificial Intelligence, ECAI 2000, Berlin, Germany, pp. 13–19 (2000)
5. Cilibrasi, R., Vitanyi, P.M.B.: Automatic meaning discovery using Google, Available at (2004), http://xxx.lanl.gov/abs/cs.CL/0412098
6. Cimiano, P., Pick, A., Schmidt, L., Staab, S.: Learning Taxonomic Relations from Heterogeneous Sources of Evidence. In: Proceedings of the ECAI 2004 Ontology Learning Workshop (2004)
7. Faatz, A., Steinmetz, R.: Ontology enrichment with texts from the WWW. In: Proceedings of Semantic Web Mining 2nd Workshop at ECML/PKDD-2000 (2002)
8. Forkner-Dunn, J.: Internet-based Patient Self-care: The Next Generation of Health Care delivery. Journal Med Internet Research 5(2) (2003)
9. Gómez-Pérez, A., Fernández-López, M., Corcho, O.: Ontological Engineering, 2nd edn. Springer, Heidelberg (2004)
10. Grefenstette, G.: SQLET: Short Query Linguistic Expansion Techniques: Palliating One-Word Queries by Providing Intermediate Structure to Text. In: Pazienza, M.T. (ed.) SCIE 1997. LNCS, vol. 1299, pp. 97–114. Springer, Heidelberg (1997)
11. Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora. In: Proceedings of the 14th International Conference on Computational Linguistics, pp. 539–545 (1992)
12. Isern, D., Moreno, A., Pedone, G., Varga, L.: An Intelligent Platform to Provide Home Care Services. In: Proceedings of the Workshop From Knowledge to Global Care. 11th Conference on Artificial Intelligence in Medicine (2007)
13. Isern, D., Sánchez, D., Moreno, A.: HeCaSe2: A Multi-Agent Ontology-Driven Guideline Enactment Engine. In: Burkhard, H.-D., Lindemann, G., Verbrugge, R., Varga, L.Z. (eds.) CEEMAS 2007. LNCS (LNAI), vol. 4696, pp. 322–324. Springer, Heidelberg (2007)
14. Kavalec, M., Maedche, A., Skátek, V.: Discovery of Lexical Entries for Non-taxonomic Relations in Ontology Learning. In: Van Emde Boas, P., Pokorný, J., Bieliková, M., Štuller, J. (eds.) SOFSEM 2004. LNCS, vol. 2932, pp. 249–256. Springer, Heidelberg (2004)
15. Khan, L., Luo, F.: Ontology Construction for Information Selection. In: Proceedings of 14th IEEE International Conference on Tools with Artificial Intelligence, pp. 122–127 (2002)
16. Lee, D., Na, J., Khoo, C.: Ontology Learning for Medical Digital Libraries. In: Sembok, T.M.T., Zaman, H.B., Chen, H., Urs, S.R., Myaeng, S.-H. (eds.) ICADL 2003. LNCS, vol. 2911, pp. 302–305. Springer, Heidelberg (2003)
17. Maedche, A., Staab, S.: Discovering Conceptual Relations from Text. In: Proceedings of the 14th European Conference on Artificial Intelligence, pp. 321–325. IOS Press, Amsterdam (2000)

18. Maedche, A., Staab, S.: Ontology Learning for the Semantic Web, IEEE Intelligent Systems. S.I. on the Semantic Web 16(2), 72–79 (2001)
19. Pasca, M.: Acquisition of Categorized Named Entities for Web Search. In: Proceedings of 13th Conference on Information and Knowledge Management, pp. 137–145 (2004)
20. Reinberger, M.L., Spyns, P.: Discovering knowledge in texts for the learning of DOGMA inspired ontologies. In: Proceedings of Workshop on Ontology Learning and Population, ECAI 2004, pp. 19–24 (2004)
21. Resnik, P., Smith, N.: The web as a parallel corpus. Computational Linguistics 29(3), 349–380 (2003)
22. Sánchez, D., Moreno, A.: Discovering Non-taxonomic Relations from the Web. In: Corchado, E.S., Yin, H., Botti, V., Fyfe, C. (eds.) IDEAL 2006. LNCS, vol. 4224, pp. 629–636. Springer, Heidelberg (2006)
23. Sánchez, D., Moreno, A.: Development of new techniques to improve Web Search. In: Proceedings of 9th International Joint Conference on Artificial Intelligence, pp. 1632–1633 (2005)
24. Sánchez, D., Moreno, A.: A methodology for knowledge acquisition from the web. International Journal of Knowledge-Based and Intelligent Engineering Systems 10(6), 453–475 (2006)
25. Sanderson, M., Croft, B.: Deriving concept hierarchies from text. In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Berkeley, USA, pp. 206–213 (1999)
26. Schutz, A., Buitelaar, P.: RelExt: A Tool for Relation Extraction in Ontology Extension. In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (eds.) ISWC 2005. LNCS, vol. 3729, pp. 593–606. Springer, Heidelberg (2005)
27. Turney, P.D.: Mining the Web for synonyms: PMI-IR versus LSA on TOEFL. In: Flach, P.A., De Raedt, L. (eds.) ECML 2001. LNCS (LNAI), vol. 2167, pp. 491–502. Springer, Heidelberg (2001)