R. Bernardi and M. Moortgat (eds.), *Questions and Answers: Theoretical and Applied Perspectives, Proceedings of the 2nd CoLogNET-Elsnet Symposium*, December 2003, Amsterdam, 5–15
http://www.let.uu.nl/~ctl/workshops/CES03/

Using ontology in query answering systems: scenarios, requirements and challenges

Werner Ceusters[a], Barry Smith[b], Maarten Van Mol[a]

[a] Language & Computing nv (L&C), Hazenakkerstraat 20a, B-9520 Zonnegem, Belgium
[b] Institute for Formal Ontology and Medical Information Science, Härtelstrasse 16-18, 04107-Leipzig, Germany

Abstract. Equipped with the ultimate query answering system, computers would finally be in a position to address all our information needs in a natural way. In this paper, we describe how Language and Computing nv (L&C), a developer of ontology-based natural language understanding systems for the healthcare domain, is working towards the ultimate Question Answering (QA) System for healthcare workers. L&C's company strategy in this area is to design in a step-by-step fashion the essential components of such a system, each component being designed to solve some one part of the total problem and at the same time reflect well-defined needs on the prat of our customers. We compare our strategy with the research roadmap proposed by the Question Answering Committee of the National Institute of Standards and Technology (NIST), paying special attention to the role of ontology.

1 Introduction

Physicians cannot be accused of lacking imagination. Since the very beginning of medical informatics they have been dreaming of computer systems that would assist them with their information needs in response to what they conceive as "simple" requests. The information needed by physicians is of two sorts. First is information concerning patients, such as the changes in Mr. X's blood pressure over the past three days, or the substances to which Mrs Y is allergic. Second is what we might call medical knowledge, i.e. the information found in textbooks, journal articles, and clinical studies - and information accumulated in the physician's brain.

It should come as no surprise that greater interest has been shown in the question of how to get such information out of the machine than how to get it in there in the first place. Entering patient related data in electronic patient record systems is a time-consuming task, not least because, as Tange *et al*. have phrased it, "*Initiatives to facilitate the entry of narrative data have focused on the control rather than the ease*

of data entry" ([1], p. 24). There are some who would see this as a minor issue, a matter of finding the right kind of clerical support. But would it not be a life saver to have in the operating theatre a QA system that, when asked whether it is safe to give the patient an additional shot of an hypotensive agent in order to reduce bleeding, would respond with: "Can you please wait for 45 seconds since the patient's blood pressure has been dropping slightly already for the last 2 minutes?"?

The research roadmap proposed by the Question Answering Committee of the National Institute of Standards and Technology (NIST) □2□ provides a good framework within which to describe the QA scene in the healthcare domain and we follow the structure of this document in what follows. NIST divides each QA-system into a question part and an answering part. Both rely on automatic reasoning strategies and processes. We first survey the kinds of questions that arise in physicians' minds in the course of their work and analyse them in terms of the research issues outlined in the NIST roadmap □2□ paying particular attention to the associated reasoning strategies. We then carry out the counterpart analysis on the answering side, indicating in each case the implications of our analyses for issues of ontology. We also discuss in light, of these analyses, the implementations which have been made in L&C's medical natural language understanding applications both in response to the needs of its customers and as a contribution to the goal of constructing the ultimate QA system for the healthcare sector.

2 The question part

2.1 What questions do physicians ask?

The answer is "many", of course, and the most important one, at least in Belgium, is "how can I make a decent, living given the abundance of doctors?" (the actual number being one doctor per 215 Belgian citizens). This however is not the kind of question we are interested in for the purposes of this paper.

Ely *et al*. [3] developed a taxonomy which comprised some 64 generic question types and used it to classify 1396 clinical questions from general practitioners and pediatricians. The three commonest generic types were: "What is the drug of choice for condition x?" (150 questions, 11%); "What is the cause of symptom x?" (115 questions, 8%); and "What test is indicated in situation x?" (112 questions, 8%). The top ten question types comprised some 64% of the total number of individual questions.

In [4], a set of 100 questions in French on oral surgery was collected, mainly from medical students, but in part also from textbooks used by students in preparing class exams in stomatology. The authors identified 66 different syntactic-semantic patterns in the questions, which they grouped into 8 classes of questions of the form:

W:(X)-(relation)-(Y),

where “W” stands in for a question operator such as ‘what?’ or ‘why?’. The top three of these classes accounted for some 90% of the questions, the first consisting of open-ended questions such as: “What diseases are evoked by the Reed-Sternberg cell?”, the second covering questions testing the validity of a proposed relationship such as: “Does X cause Y?”, the third asking for explanations as in: “Why does X cause Y?”.

Another source of medically relevant questions can be found in the OHSUMED corpus [5] of queries entered into an information retrieval system. Novice physicians using MEDLINE generated 106 queries, having been asked beforehand to provide information about the relevant patient and about their own information needs. The questions in this collection are not standard grammatical questions but rather have forms such as: “60 year old with lung abscess, surgery vs. percutaneous drainage for lung abscess”, which could be rephrased as “Should a 60 year old patient with lung abscess be treated with surgery or percutaneous drainage?”.

Studies like the ones mentioned above give an indication both of how easy it might at first sight seem to design a system that can understand the different types of questions and also of how difficult it is to realize this task if one wants to take every parameter into account. But from an industrial perspective a centrally important question is: how many of those questions generated during actual clinical encounters or interventions at the point of care could not be answered using traditional methods such as asking a colleague or consulting a clinical guideline or an Internet search engine. From an industrial point of view one must consider also issues such as whether a right answer in fact exists given that many relevant medical questions remain still unanswered, whether an existing source known as being useful is available on the spot, or whether there was enough time to carry out an adequate search. Ebel and White showed that 0.42 unanswered questions were generated per patient-physician contact, of which the majority was rated to be sufficiently important that an answer might have led to different advice or treatment [6]. Combine this with studies which show that in the United States between 44,000 and 98,000 people are killed every year from medical errors and that the total cost of preventable medical mistakes, including lost wages and extra health costs, are estimated to lie between $17 billion and $29 billion a year [7], and it becomes obvious that good QA systems reflect a pressing need.

2.2 Ontology for question analysis

The range of clinical questions covers the entire spectrum from very simple to the very complex spectrum as described in [8], with the majority appearing (superficially at least) to be of the most simple type. Although a question such as “What is the best therapy for disease X?” contains a judgement request and thus is normally considered to be complex, the answer can often be found in clinical guidelines in which judgements from authoritative sources have been laid down a priori. In consequence, such questions may be labeled “simple factual”.

Studies like the ones mentioned above show that healthcare informatics research accepts the need for question taxonomies along the lines proposed in [8], though the issue is addressed from a number of different perspectives, reflecting opportunities for cross-disciplinary fertilisation. Such studies have also started to address the second

issue related to the question part of a QA system, namely the need for "*a semantic model of question understanding and processing, one that would recognize equivalent questions, regardless of the speech act or of the words, syntactic inter-relations or idiomatic forms"* [2, p. 8].

What is striking however is that no one has thus far attempted to bring the typology of questions mentioned above into some direct relation to the individual patient. One may congratulate a student who gives the best possible answer to the question "Should a 60 year old patient with lung abscess be treated with surgery or percutaneous drainage?", but in actual practice the more appropriate question would be: "May I give THIS 60 year old patient with THIS lung abscess surgery or percutaneous drainage?". [8] correctly recognizes that ontology can serve as a means for dealing with the actual context in which questions arise. What it does not recognize, however, is that for this purpose one requires an ontological theory that comprehends not only *classes* but also individual *instances,* i.e. entities that are bound to specific (normally topologically connected) locations in space and time [9, 10]. The ontology management system and servers we require must thus be able to cope with this requirement, and this in spite of the fact that we are living in an era where massive interest in what Brachman called the T-Box (of concepts) [11] seems to have caused most researchers to forget about the A-Box (the individual instances in spatio-temporal reality).

2.3 Question analysis technology at L&C

With respect to the question part of a QA system, L&C currently has no dedicated question analysis software devoted to grammatical questions formulated in natural language. Query analysis in L&C's information retrieval system rather is a special case of a general methodology of document analysis which we now described. The technologies currently used in realizing this methodology are LinkFactory®/LinK-Base® for ontology maintenance and use, and MaDBoKS for linking LinKBase® to instance databases.

2.3.1 LinkFactory®/LinKBase®

LinKBase® is a large-scale medical ontology developed by L&C using the ontology authoring environment LinKFactory® [12]. LinKBase® contains over 2 million language-independent medical and general-purpose concepts, which are associated with more than 4 million terms derived from a number of different natural language sources [13]. A *term* consists of one or more *words*, which may be associated with other concepts in their turn. Concepts are linked together into a semantic network in which some 480 different link types are used to express different sorts of relationships. The latter are derived from formal-ontological theories of mereology and topology [14, 15], time and causality [16], and also from the specific requirements of semantics-driven natural language understanding [17, 18]. Link types form a multi-parented hierarchy in their own right. At the heart of this network is the formal subsumption (is-a) relationship, which in LinKBase® covers only some 15% of the total number of relationships involved. Currently, the system is being re-

engineered in conformity with the theories of Granular Partitions [19] and Basic Formal Ontology [10, 20].

2.3.2 MaDBoKS

The MaDBoKS (Mapping Databases onto Knowledge Systems) tool is an extension of the Linkfactory® to link external relational databases to LinkBase® [21]. Database schemas from existing databases, for example from hospital patient databases or electronic patient records, can be retrieved and mapped to the ontology in such a way as to establish a two-way communication between each of the databases and the ontology. The ontology thereby serves as a central switchboard for data integration and the database schemas themselves thus come to function as semantic representations of the underlying data (analogous to the semantic representations of natural language utterances that are processed by natural language understanding software). In a natural language understanding system, a semantic parser bridges the gap between the ontology and the documents from which information is to be extracted. Here an analogous piece of software, called a *mediator*, bridges the gap between the ontology and the databases to be integrated.

3 The answer component of a clinical QA system

3.1 Data sources

There is of course in the healthcare domain no shortage of information sources from which appropriate answers can be derived. MEDLINE® (Medical Literature, Analysis, and Retrieval System Online) is the primary bibliographic database of the U.S. National Library of Medicine (NLM), containing over 12 million references to journal articles in the life sciences with a concentration on biomedicine. It can be searched via PubMed® or the NLM Gateway and covers 1966 to the present. Citations come from over 4,600 journals in 30 languages [22]. There are in addition a multitude of Web portals oriented towards either health professionals or consumers. But this does not of course mean that it is easy to find the right answers to your questions, and this is so whether the answer is sought by an automated process or through the mediation of a skilled professional. Ely *et al*. [23] report that the major obstacles for physicians with respect to their information needs include:

- • difficulties in selecting an optimal strategy to search for information (deciding which resources will be most helpful; deciding in which order to search; which articles to read thoroughly and how thoroughly, and so forth);
- • uncertainty in establishing when all relevant evidence has been found so that the search can stop;
- • inadequacies in the synthesis of disparate bits of evidence into a clinically useful statement (including problems raised by conflicting evidence);

- • the real or apparent absence of appropriate resources to cover a given topic, even in spite of the overwhelming mass of information that is in principle available.

Not surprisingly, these obstacles correspond exactly to the challenges facing answer extraction systems mentioned in the research roadmap document [2].

3.2 Ontology for answer processing

The latter mentions also the need for using ontology for purposes of answer processing, though it leaves us unclear as to what the authors precisely understand by this term.

The medical informatics field is blessed not only with a wealth of data sources but also with many lexical resources, thesauri, terminology systems and ontologies and all of these, despite problems related to quality [24, 25], can be used for query answering purposes, whether as sources in their own right or as tools to be used e.g. in the processing of natural language documents. Currently, they are overwhelmingly used for document retrieval, a very important component in question answering systems, the state of the art being described in [26].

3.3 L&C technology for the answer component of QA systems

L&C has developed a number of applications that are essential components of an answering system, but not yet everything that is required to have a fully functioning QA system.

3.3.1 TeSSI®: Terminology Supported Semantic Indexing

TeSSI® is a software application performing semantic indexing. TeSSI® first segments a document into its individual words and phrases. It then matches words and phrases in the document to corresponding LinKBase® entities using a semantic lexicon [27]. This step introduces ambiguity, since some entities share homonymous terms. To resolve cases of ambiguity, TeSSI® uses domain knowledge from LinkBase® to identify which concept out of the set of concepts that are linked to a homonymous phrase best fits with the meaning of the surrounding terms in the document.

In the next step, TeSSI® uses the matches between concepts identified in the document and the domain knowledge in LinkBase® to infer additional concepts which are only implicitly part of the subject matter of the document. The end result of this process is a graph structure in which the nodes correspond to the LinkBase® entities explicitly or implicitly present in the document, and in which these nodes are joined by two kinds of arc corresponding respectively to semantic relationships derived from the LinkBase® domain ontology and to co-occurrence relationships derived from the position of terms in the document, the inclusion of the latter being motivated by many studies showing that co-occurring terms are often semantically related [28]. The arcs are weighted according to the semantic distance between the corresponding entities in LinKBase® and according to the proximity of the

corresponding terms in the document. The nodes are weighted based on the number of occurrences of the corresponding terms in the document.

Having identified all the medical (and non-medical) concepts in a document, TeSSI® then ranks these concepts in the order of their relevance to the document as a whole, hence identifying the topics of the document. Relevance scores are on a scale of 0 to 100, with 100 representing the most relevant concept. To determine these scores, TeSSI® uses a constraint spreading activation algorithm on the constructed graph [29]. In this way, semantically related concepts reinforce each other's relevance rankings. The rationale for this algorithm stems from the observation that the concepts in any particular document will vary in their degree of semantic independence from each other. For example, a document might contain one mention each of the terms "heart failure," "aortic stenosis," and "headache", the first two of which are clearly more closely related to each other than to the third. An indexing system based entirely on term or concept frequency will treat these three concepts independently, thus assigning them all the same relevance. Intuitively, however, the document has twice as many mentions of heart disease as of headache. TeSSI® takes advantage of its underlying medical ontology in order to represent more accurately this type of phenomenon.

3.3.2 L&C's Information Extraction System

The L&C Information Extraction system consists of a number of components that successively add structured information to an unstructured text. Some of these components are necessary elements of the system, others are optional. The system takes a text document in natural language as input and creates an XML document as output. The latter then serves as the basis for further user-defined operations including querying and template-filling. As such, the information extraction system itself is an essential component in a query answering system. The XML output is created via Natural Language Processing (NLP) involving Full Syntactic Parsing for syntactic analysis and LinKBase® for analysing semantics. In addition Text Grammar Analysis (TGA) is applied, which means that the system looks for relations between parts of text. We believe that it is only through TGA carried out on top of syntactic and semantic analysis of individual sentences that a full understanding of the meaning of text in natural language will be possible in the future. We will now describe the details of the resulting system.

The basic components of the L&C information extraction system are:

- Segmentation
- Section Labeling
- Clause/Phrase Segmentation
- Fragment Labeling
- Information Extraction,

and we deal with each of these in turn.

The input text is first segmented into paragraphs and sentences. Each sentence is then decomposed into its basic constituents, which are then tagged with markers for syntactic and semantic information. Segmentation uses rules easily adaptable to the

client's particular document requirements. An important step in the process is *section segmentation* carried out at document level rather than sentence level. A text is not an unordered succession of a number of data, but rather a structured whole in which all information comes at a certain functional stage. Recognizing the different sections in a text is thus important for getting at the meaning of a text. As an example, the first sections of this paper are: title, authors and affiliation, and abstract. In medical discharge summaries, typical sections are: patient-related administrative data, anamnesis, clinical findings, and so forth.

Each section is then automatically labeled with a label that reflects its meaning. Labeled sections are used to limit the scope of search when looking for particular information to be extracted. For example, discharge medication will only be looked for in sections in which discharge medication is known to appear. Labeling is based on labeling rules, gathered from a training corpus, that take into account a number of weighted features. Clients of our system can choose whether they want to use an existing training corpus or to create their own. If they choose the latter they are supplied with a fully customizable basic set of possible section labels with their descriptions. A user-friendly Graphical User Interface for labeling texts is included with the system. It can be used to first label manually a training corpus that then serves as input for a supervized learning algorithm that generates the rules to label similar texts automatically. Labeled sections are used to limit the scope of search when looking for particular information to be extracted. The accuracy of the labeler for medical discharge summaries was shown in 2001 to be 96.4% [30] and amounts currently to 97.23% (tested on 4421 sections in 100 medical reports).

Each sentence can be further subdivided into clauses and phrases. To do this we use our Full Syntactic Parser, a hybrid system combining both symbolic and statistical approaches. We use a dependency grammar-based formalism to capture the syntactic relations between the words in a sentence. Using a dependency formalism enables us for instance to immediately capture the scope of negated elements in a sentence. The grammar is developed in a graphical environment, enabling us to develop grammars rapidly with substantial coverage.

The different fragments that are recognized by the Clause/Phrase Segmenter with embedded Full Syntactic Parser receive a functional label - such as "clinical finding" or "diagnosis" - according to their content. The Fragment Labeler uses the same techniques as the Section Labeler and thus also needs to be built up by means of a training corpus, which again can be provided either by L&C or by the client. Fragment label information is used to further narrow down the amount of text in which information will be searched for.

The Information Extraction component uses information from the Section Labeler and the Fragment Labeler, as well as conceptual information from LinKBase®, syntactic information from our Full Syntactic Parser, and novel machine learning techniques which in combination go much further than standard text analysis algorithms which rely on string matching and similar techniques.

4 Discussion and conclusion

Whether ontology can help to build better QA systems in the healthcare domain is an issue that has so far not been proven, though there is considerable scientific evidence to suggest that the answer will be positive. Our personal experience in using LinkBase® in a variety of natural language understanding components suggests a partial proof of this thesis, and in a way which gives L&C a competitive advantage [31] as recognized for example by the technology watchers Frost and Sullivan, who presented L&C with the Healthcare Information Technology & Life Sciences Product of the Year Award at the 2003 Global Excellence in Healthcare & Life Sciences Awards Banquet in San Diego, November 6th [32]. On the other hand, the costs involved in developing good ontologies are considerable, so that it is no surprise that some argue that less powerful terminology-based systems would be sufficient for purposes such as document retrieval. Thus it is a relief that this is no longer the vision of (people inside) the National Library of Medicine. As Olivier Bodenreider states:

> The UMLS is an extensive source of biomedical concepts. It also provides a large number of inter-concept relationships and qualifies for a source of semantic spaces in the biomedical domain. However, the organization of knowledge in the UMLS is not principled nor consistent enough for it to qualify as an ontology of the biomedical domain. In the tradition of the UMLS, the approach we propose for going toward an ontology consists of refining the definition and organization of the existing semantic space. Both basic and applied research is needed to augment and better organize knowledge in the UMLS. A sound, ontological representation of biomedical knowledge is expected to enable tasks such as reasoning, currently hardly possible with the UMLS, while improving the performance of tasks already supported (e.g., information retrieval). [33, p. 22].

Since the task to be performed by QA systems is extremely complex, one can expect the supporting ontologies to be equally complex, covering not only domain knowledge, but also areas such as user beliefs and expectations. We strongly believe that only an ontological theory grounded in realism and a focus on both classes and instances can meet these demands, and that such an ontological theory is a necessary perquisite of the query-answering systems of the future.

5 References

1. Tange HJ, Hasman A, de Vries Robbe PF, Schouten HC. Medical narratives in electronic medical records. International Journal of Medical Informatics, 1997, 46(1): 7-29.
2. Burger, John; Cardie, Claire; Chaudhri, Vinay; Gaizauskas, Robert; Harabagiu, Sanda; Israel, David; Jacquemin, Christian; Lin, Chin-Yew; Maiorano, Steve; Miller, George; Moldovan, Dan; Ogden, Bill; Prager, John; Riloff, Ellen; Singhal, Amit; Shrihari, Rohini; Strzalkowski, Tomek; Voorhees, Ellen; Weischedel, Ralph. (2002) “Issues, Tasks and

Program Structures to Roadmap Research in Question & Answering (Q&A). www-nlpir.nist.gov/ projects/duc/papers/qa.Roadmap-paper_v2.doc
3. John W Ely, Jerome A Osheroff, Paul N Gorman, Mark H Ebell, M Lee Chambliss, Eric A Pifer, and P Zoe Stavri. A taxonomy of generic clinical questions: classification study. BMJ 2000;321 429-432.
4. Pierre Jacquemart and Pierre Zweigenbaum. 2003. Towards a medical question-answering system: a feasibility study. In Pierre Le Beux and Robert Baud, editors,*Proceedings Medical Informatics Europe*, Amsterdam. IOS Press 2003; 463-468.
5. W. R. Hersh, C. Buckley, T. J. Leone, and D. H. Hickam. OHSUMED: an interactive retrieval evaluation and new large test collection for research. In: Proceedings of the 17th Annual International ACM Special Interest Group in Information Retrieval. Dublin, Ireland: Springer-Verlag, 1994:192--201.
6. Mark H. Ebell and Linda White. What is the best way to gather clinical questions from physicians? J Med Libr Assoc. 2003 July; 91 (3): 364-366.
7. Institute of Medicine "To Err Is Human: Building A Safer Health System.". December 1999.
8. Jaime Carbonell, Donna Harman, Eduard Hovy, Steve Maiorano, John Prange,Karen Sparck-Jones, Vision Statement to Guide Research in Question & Answering (Q&A) and Text Summarization. http://www-nlpir.nist.gov/projects/duc/papers/Final-Vision-Paper-v1a.doc.
9. Thomas Bittner and Barry Smith. Directly Depicting Granular Ontologies; presented at the 1st International Workshop on Adaptive Multimedia Retrieval, Hamburg, September 2003 (http://wings.buffalo.edu/philosophy/faculty/smith/articles/DDGO.pdf).
10. Grenon P, Smith B. SNAP and SPAN: Prolegomenon to geodynamic ontology, forthcoming in Spatial Cognition and Computation.
11. Brachman R, “On the Epistemological Status of Semantic Networks,” In Findler,N. (ed.). *Associative Networks: Representation and Use of Knowledge by Computers,* Academic Press, New York, 1979; 3-50.
12. Ceusters W, Martens P, Dhaen C, Terzic B. LinKBase: an Advanced Formal Ontology Management System. Interactive Tools for Knowledge Capture Workshop, KCAP-2001, October 2001, Victoria B.C., Canada (http://sern.ucalgary.ca/ksi/K-CAP/K-CAP2001/).
13. Montyne F, The importance of formal ontologies: a case study in occupational health. OES-SEO2001 International Workshop on Open Enterprise Solutions: Systems, Experiences, and Organizations, Rome, September 2001 (http://cersi.luiss.it/oesseo2001/papers/28.pdf).
15. Smith B. Mereotopology: a theory of parts and boundaries, Data and Knowledge Engineering 1996; 20: 287-301.
14. Smith B, Varzi AC. Fiat and bona fide boundaries, Proc COSIT-97, Berlin: Springer. 1997: 103-119.
16. Buekens F, Ceusters W, De Moor G. The explanatory role of events in causal and temporal reasoning in medicine. Met Inform Med 1993; 32: 274-278.
17. Ceusters W, Buekens F, De Moor G, Waagmeester A. The distinction between linguistic and conceptual semantics in medical terminology and its implications for NLP-based knowledge acquisition. Met Inform Med 1998; 37(4/5): 327-33.
18. Bateman JA. Ontology construction and natural language. Proc Int Workshop on Formal Ontology. Padua, Italy, 1993: 83-93.
19. Bittner T, Smith B. A theory of granular partitions. *Foundations of Geographic Information Science*, Duckham M, Goodchild MF and Worboys MF, eds., London: Taylor & Francis Books, 2003: 117-151.
20. James M. Fielding, Jonathan Simon, Barry Smith. Formal Ontology for Biomedical Knowledge Systems Integration. Submitted to MEDINFO2004. (http://ontology.buffalo.edu/medo/FOBKSI.pdf).

21. Verschelde JL, Casella Dos Santos M, Deray T, Smith B and Ceusters W. Ontology-assisted database integration to support natural language processing and biomedical data-mining, Journal of Integrative Bioinformatics 2003 (submitted).
22. National Library of Medicine. Fact Sheet Medline® (http://www.nlm.nih.gov/pubs/factsheets/medline.html)
23. Ely JW, Osheroff JA, Ebell MH, Chambliss ML, Vinson DC, Stevermer JJ, Pifer EA. Obstacles to answering doctors' questions about patient care with evidence: qualitative study. BMJ 2002;324:710-713.
24. Ceusters W, Smith B. Ontology and Medical Terminology: why Descriptions Logics are not enough. TEPR 2003 (electronic publication).
25. Ceusters W, Smith B, Kumar A, Dhaen C. Mistakes in Medical Ontologies: Where Do They Come From and How Can They Be Detected? in Pisanelli DM (ed) "Ontologies in Medicine. Proceedings of the Workshop on Medical Ontologies, Rome October 2003" IOS Press, 2004 (in press).
26. William Hersh. Information Retrieval: A Health and Biomedical Perspective (Second Edition). Springer Verlag, 2003.
27. Jackson B, Ceusters W. *A novel approach to semantic indexing combining ontology-based semantic weights and in-document concept co-occurrences.* In Baud R, Ruch P. (eds) EFMI Workshop on Natural Language Processing in Biomedical Applications, 8-9 March, 2002, Cyprus, 75-80.
28. Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. Journal of the American Society for Information Science, 41(6), 391-407
29. J.A. Hendler, Marker-Passing over Microfeatures: Towards a Hybrid Symbolic/Connectionist Model. Cognitive Science 1989 (1) 79-106.
30. Maarten Van Mol & Mick O'Donnell. Automatic Recognition of Generic Structure: Medical Discharge Notices. In: Text and Texture, Systemic Functional viewpoints on the nature and structure of text. L'Harmattan, Paris, 2004 (in press).
31. Smith B, Ceusters W. Towards Industrial-Strength Philosophy; How Analytical Ontology Can Help Medical Informatics. Interdisciplinary Science Reviews, 2003, vol 28, no 2, 106-111.
32. http://www.chron.com/cs/CDA/story.hts/prn/texas/2205626
33. Olivier Bodenreider. Medical Ontology Research. A Report to the Board of Scientific Counselors of the Lister Hill National Center for Biomedical Communications, May 17, 2001.