

Context Related Extraction of Conceptual Information from Electronic Health Records

Svetla Boytcheva¹, Ivelina Nikolova² and Elena Paskaleva²

¹ State University of Library Studies and Information Technologies
119 Tzarigradsko Shose Blvd., 1784 Sofia, Bulgaria,
svetla.boytcheva@gmail.com

² Institute for Parallel Processing, Bulgarian Academy of Sciences
25A Acad. G. Bonchev Str., 1113 Sofia, Bulgaria,
iva@lml.bas.bg helen@lml.bas.bg

Abstract. This paper discusses some language technologies applied for the automatic processing of Electronic Health Records in Bulgarian, in order to extract multi-layer conceptual chunks from medical texts. We consider an Information Extraction view to text processing, where semantic information is extracted using predefined templates. At the first step the templates are filled in with information about the patient status. Afterwards the system excerpts or infers temporal relations between the events, described in the EHR text. Then cause-effect relations are explicated and at last, implicit knowledge is derived from the medical records after reasoning. Thus we propose a cascade approach for the extraction of multi-layer knowledge representation statements because the subject is too complex. In this article we present laboratory prototypes for the first two tasks and discuss typical examples of conceptual structures, which cover the most challenging tasks in the extraction scenario - the recognition of cause-effect relations and temporal structures. The present work in progress is part of the research project EVTIMA (2009-2011) that aims at the design and implementation of technologies for efficient search of conceptual patterns in medical information.

1 Introduction

The case history is one of the most important medical documents that were created, processed and stored since the ancient times. Medical patient records enable doing research on disease causes and disease symptoms as well as searching for effective treatment methods. Unfortunately, most of these medical documents in Bulgaria are available as paper archives or in digital format as separate text files only.

The needs for more precise clinical research on disease reasons and their prevention, as well as for their effective treatment gave rise to large repositories of Electronic Health Records (EHRs) which can be explored by computers. Modern medical informatics requires the development of effective methods for conceptual information extraction from text documents and creation of EHR

data bases in appropriate format, thus facilitating the application of advanced methods and techniques for knowledge discovery in medical documents. However the unstructured text of the medical records and the various ways used to refer to the same medical condition (e.g. disease, symptom, examination results) make the automated analysis a challenging task.

There are different types of Electronic Medical Records - one managed by the patient's GP, others with epicrisis issued by hospitals and so on. Here we shall work with EHRs provided for hospital treatments. Due to the particular requirements for the personal data protection and restricted access to the medical documents, the project works on anonymous patient data. The pseudonymisation is done by the Hospital Information System of the University Specialised Hospital for Active Treatment of Endocrinology Acad. I. Penchev (part of the Medical University, Sofia). The project deals with the automatic processing of epicrisis of patients who are diagnosticised with different forms of diabetes.

The paper is structured as follows. Section 2 overviews some related research and discusses basic language technologies which are used for Information Extraction (IE) in the medical domain. Section 3 describes the raw data features and an outline of the data processing towards knowledge acquisition. Section 4 discusses the main types of patient status data and some techniques for their extraction. Section 5 presents the module for extraction of temporal structures from EHRs. Section 6 briefly sketches the ideas and techniques behind the modules for cause-effect relation detection and more complex relations inferences. Examples and assessment figures present the current experiments. Section 7 contains some discussions and the conclusion.

2 Related work

The main approach for partial text processing is called Information Extraction (IE). The goal of the IE systems is to search for a given event in the input documents [1]. Different IE systems for various domains have been developed and they process text in different natural languages. The last two decades Natural Language Processing (NLP) methods have started to penetrate the medical field and provide the extraction of medical entities and their classification. Today medical IE is integrated in various kinds of applications, for instance entity extraction from patient records starting from some initial medical ontology [2] or study of large medical databases to identify diseases which co-occur together more commonly [3]. The most challenging task is free text processing in the medical domain. For the latter techniques are used such as:

- matching existing domain ontology to patient record texts [4–6] - knowledge-based IE is a very difficult task since the conceptual resources are very often incomplete, so important terms from the EHR text might be missing in the ontology. Moreover the concept labels are usually presented in the domain ontologies in their canonical form only. These systems are not very successful in the recognition of paraphrases, compound concepts, and concepts that incorporate critical modifiers;

- rule-based processing [7, 8] - usually these systems have drawbacks like (i) difficulties in generalisation/specialisation of the manually prepared rules, (ii) rule set management and consistency support and (iii) adaptation of the rule set for another subdomain;
- phrase spotting [9] - these systems search for specific key terms or phrases in the medical records but often they fail to recognise paraphrases and compound concepts. Another problem is that the particular context of the key term can differ from the required one, moreover the key term can also occur in a negated context [10];
- mapping the input text to an internal language-independent representation, like in the systems RECIT [11] and GMT [12] - this approach also has limitations in the analysis of complex language constructions.

More generally, semantic interpretation of medical data can be done by encoding knowledge structures as Conceptual Graphs (CG). Croitoru et al [13] present a CG approach to describing data in a distributed decision support system for brain tumour diagnosis. They use CG for capturing the static model rather than the inference procedures. Specific CG extensions like the notion of "actor" prove to be useful because they provide higher-level functionality on top of the knowledge statements. In [14] Delugach proposes to model multiple medical views by CG; he introduces "actors" to represent some relationships in the data flow diagrams for medical information. The "Actors" in [15] are applied to automate a requirements consistency checks by CG representation; this idea is useful in the medical domain too because of the many requirements. Pfeiffer and Hartley [16] present a Conceptual Programming environment, where graph definitions contain relational functions, a kind of "actors", for allowing state, event, and temporal processing.

Some IE systems in the medical domain are considered as best practices, for instance BADGER [17] - a text analysis system, which summarizes medical patient records by extracting concepts (diagnoses, symptoms, physical findings, test results, and therapeutic treatments) based on linguistic context and AMBIT [18] - a system for information extraction from biomedical texts.

3 Project settings - from data to knowledge

In Bulgaria the EHR content is standardised by state regulatory documents. The length of EHR text is usually 2-3 pages and the text is separated into the following parts according to the medical standards: General information for the patient, Diagnosis, Anamnesis, Accompanying/Past diseases; Family anamnesis; Risk factors, Status, Examinations and Clinical Data; Consultations; Debate.

One of the major problems in processing Bulgarian EHRs is the variety of terminological expressions. The text contains medical terminology in Latin (about 1%), sometimes with different transcriptions in Cyrillic. Several wordforms per term are used for most of the medical terminology in Bulgarian (66%) and there are specific abbreviations of the medical terminology both in Bulgarian and Latin (about 3%). The EHRs contain also a lot of numerical values of analyses and

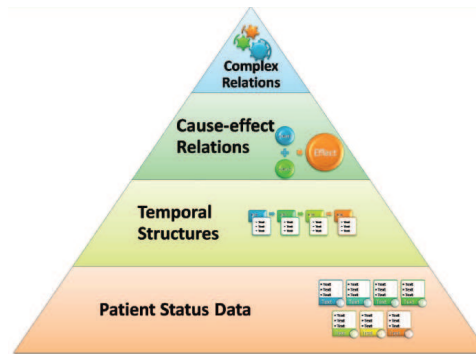


Fig. 1. Knowledge Representation Hierarchy

clinical test data (about 16%). Another problem is the specific language style of the medical professionals. The major part of the text consists of sentence phrases; complete sentences are rare. Sometimes there is no agreement between the sentence parts. All these obstacles together are not easy to overcome. They prevent the effective application of standard NLP techniques like deep syntax analysis. Moreover, especially for Bulgarian, there is a lack of well-developed, stable language technologies with satisfactory precision for partial syntax analysis. Another essential problem is the rich temporal structure of the patient descriptions - about 2/3 of the whole text describe connected events, which have happened in different periods of time. Due to all reasons listed above, it is necessary to elaborate NLP techniques for conceptual structures extraction using partial analysis of the Bulgarian EHRs.

The desired knowledge structures to be extracted from EHRs can be grouped in the following types: *(i)* Characteristics for filling information in templates for patient status, diagnoses, symptoms, treatment etc.; *(ii)* Discovering relations - cause-effect relations between symptoms-diagnosis-treatment etc.; *(iii)* Analysis of the different treatments effect on the patient status.

This information is represented in different levels of abstraction, therefore it is necessary to construct a multi-layered knowledge representation hierarchy (Fig. 1). The lowest level consists of data about the patient status - symptoms, clinical data and diagnosis. At this level we use IE techniques for filling in templates with information about body parts and their current status, symptoms and/or diagnosis; IE operates on the text words. The middle level represents temporal relations between events (data and events' order). The next level shows cause-effect relations (meaning of data), and the uppermost level corresponds to implicit knowledge extracted from medical records after reasoning (understanding the data). Thus there are several challenging tasks in solving complex relations recognition: filling in templates with patient's data, temporal structures determination (by extraction of sequences of events) and recognition of the cause-effect paradigm.

4 Patient Status Data Extraction

Patient-related documents show significant deviation from the regular text where coherent discourse is built by adjacent sentences. The EHR texts are even more specific case, because they are a kind of non-official, non-edited hospital documentation, which is entered by different authors. These observations are supported by statistical analyses of the experimental corpus tokens.

The primary, raw text corpus consists of 106 EHRs forming a corpus of 73 600 word occurrences. After the elimination of repeated identification records and long digit expressions (tables, enumerations etc.) the corpus was reduced to 65 600 tokens. 10% of them are non-language units - Latin words, digits and mixed letter-digit strings. The rest of the corpus - 58 920 Bulgarian words were annotated with a large Bulgarian grammatical dictionary (1 100 000 wordforms of 75 000 lemmas). The results of the morphosyntactical annotation showed a large percentage of unrecognised words (13 370 tokens or 22,6% of all running Bulgarian words). The 45 550 analysed tokens are presented by 2 600 wordforms and their 1 840 lemmas.

The results given in Fig. 2 show the high percentage of extra-language and extra-dictionary data in the medical texts (total - 30%). For comparison, the correlation of such data in narrative texts is given in Fig .2 as well (without extra-language units and only 5,8% extra-dictionary units). The figures show also the "poverty of expressions" in medical texts - the list of different wordforms - 5,7% of the running words, while the same value for narrative texts is 19,4%.

	Tokens total	Non-language units	Non-language (%)	Non-dictionary units	Non-dictionary (%)	Recognised tokens	Word-forms	Word-forms (%)	Lemmas
Medical									
Plain corpus	65 600	6 680	10,2	13 370	20,3	45 550			
Recognised						45 550	2600	5,7	1 840
Narrative									
Plain corpus	57 600	0	0	3 396	5,8				
Recognised	54204					54 204	10 204	19,4	1 647

Fig. 2. Medical vs. narrative corpus - correlation between tokens, language units, extra-language units, extra-dictionary units (not found), wordforms and lemmas.

The specific text features prevent the extraction of conceptual information by the classical deep NLP techniques (e.g., tagging and deep parsing), because the key elements (for the text analysis and contents) are coded by non-language means. The usual approach in such cases is to use statistical methods for text analysis. But these methods require the presence of a large text corpus for observation, training and testing, which does not exist for Bulgarian EHRs. That is why we apply only partial chunking for extraction of Noun Phrases (NPs) wherever possible and reduce the wordforms of the extracted NPs to common

stems using stemming rules [21]. Later on, these NPs and diseases' definitions given in the International Classification of Diseases (ICD-10) are treated as key phrases for further language analysis of the patient EHRs. We build templates to recognise some crucial elements for the text comprehension. By filling these templates we obtain information about body parts and their current status, symptoms and/or diagnosis in expressions like:

- *diagnosis*: "Захарен диабет тип 1" (diabetes type 1);
- *risk factors*: "пушач" (smoker);
- *body parts*: "много суха кожа" (very dry skin); "женски тип окосмяване"; "крайници - без отоци".

Bulgarian language with its rich morphological structure enables the application of unsupervised knowledge extraction methods. Here we discuss some linguistic techniques and language oriented surface rules, which provide the construction of semantic links in the template of risk factors.

Analysing the descriptions of risk factors in 106 EHR texts, we find some linguistic dependencies between the lexical units and the concepts they represent. The communicative role of that text fragment is to serve as a link between the disease and its originating and aggravating factors (drinking, smoking) as well as between the medication and the factors for its failure or complications (e.g. allergies). A main risk factor such as smoking can be expressed by the verb пуша (smoking) as well as by its related words - the nouns пушач (smoker), пушене (smoking), тютюнопушене (tabaco-smoking). In this example the common semantic element "smoke" can be extracted from all stems of the word family so the stemming facilitates the meaning extraction from the EHR.

Besides the morphologically related words, the text expressions of the same risk factor can be realised through semantically related words like 'cigarettes' and 'box', used as a measure for the risk level. This level is normally expressed in numbers, accompanying the focal words (e.g. 2-3 cigarettes vs. 2-3 boxes per day). We code explicitly the relation between the risk action and the measurable units of the observed action. The elements of this relation are obtained by bottom-up study of the possible semantic relations in a representative text corpus. For example the risk factor 'alcohol drinking' is registered in the EHRs as "uses 100 g concentrate a day", which presumes a priori corelations between the notions of 'concentrate' and 'drinking' (even when the second word is missing in the particular text).

When building the semantic templates, special attention is given to the various language means of expressing negation [19]. Negation recognition methods should work at text level, because the negated element (keyword) might be extracted wrongly if the negation scope goes beyond the standard negative interpretation of immediate adjacency (X - no X, not X). Standard negation links denoting the absence of risk are expressed by verbs as 'отрича' (deny) , 'не съобщава' (does not notify), 'не са забелязани' (are not noticed).

Example: Fig. 3 presents a sample template for detecting the risk factor "smoking" build on the basis of all epicrisis in our corpora. Using concordances

Availability of risk factor	Degree	Quantity	Negation
пушач (smoker)	умерено (moderately)	1 (2) кутии (box/es)	не пуши (doesn't smoke)
пуши (smokes)	рядко (rarely)	1 (2) цигара/и (cigarette/s)	бивш пушач (ex smoker)
пушене (smoking)			отрича (denies)
тютюнопушене (tabaco-smoking)			

Fig. 3. Template slots and possible values for the risk factor "smoker"

and partial chunking we have collected semi-automatically the linguistic features which signify the presence, lack or frequency of smoking. Due to the impossibility to perform a deep syntactico-semantic analysis of the epicrisis text (and to obtain the detailed structure of the connected text units), the template in Fig. 3 is applied as a specific search window. Further analysis of EHRs is implemented by shallow method which presumes the definition of text limits in the search procedure - a text fragment between two punctuation marks.

5 Temporal Structures Extraction

Temporal events can be grouped in the following types: diagnosis, symptoms, and treatments. The events features are: start of the event, end of the event, type, characteristic, effect.

The Temporal Structures Extraction Module has a pipeline architecture including submodules for: Annotation Analyses & Chunking; Temporal Information Extraction and Filling Templates. This module processes the EHRs separately one by one.

The first submodule splits the EHRs into their main sub-sections. After that the sections are grouped into five major segments according to the tense of the events that will be searched for in the segment: past events, events at the moment of entering the hospital, events at the moment of leaving the hospital, continuous events, events in the future.

If some EHR part contains information about events that can be associated with more than one tense type, then that EHR part is duplicated and recorded into the corresponding tense segments. After that the annotation submodule starts the morphological analysis based on a large dictionary (described in section 4) and grammar rules. For each wordform, the module finds its basic form (lexeme) with the associated lexical and grammatical features. The lexicon is extended with a terminological bank for medical terms, derived from the ICD-10 in Bulgarian. This additional recourse contains 10970 terms, partial taxonomy of body parts, and a medical goods list.

The annotation submodule performs also chunking of the main sentence phrases, which is a partial syntactic analysis, and recognizes the noun phrases.

Each sentence in every segment is numbered separately following the sentence occurrences in EHR text.

The second submodule extracts temporal structures from the text by using keywords to locate the necessary information and rules for determining the scope of the event.

The keywords are grouped in the following categories: for time/date, events with alternating and/or undefined time, treatment events, diagnosis events, symptom events, effect.

The extraction of the temporal features is done by application of rules for discovering the temporal structures. They work on words and their annotations, inserted in the text by the partial syntax analysis module, and by determining of the category of the events (past, present, future, continuous) in the sentences which contain the corresponding "keywords". The next step is to partially order the events according to their start/end times. The system tries to find in the medical archive previous patient records (if any), to build a complete picture about the past patient case history and the newly generated events. The medical records resource bank contains patients' information in XML format. The submodule for template filling determines the scope of each event. Templates are manually generated after joint discussions with medical experts. One sentence might contain information about sequence of events but it is also possible to have the information about one event spread out among several adjacent sentences. Due to this reason, to determine the scope of the event X, the module analyses all events generated from the sentence where X is described and forms its previous and next sentences, i.e. the events ordering module processes the discourse structure of the text [19]. For each event the module determines its characteristics and effect (if any). During the process of templates filling it is necessary to define correlations between the events, depending on their sequence. The final step is to update the patient information in the resource bank with medical records.

Example for Filling of Temporal Conceptual Templates

Let an EHR for some patient be given. We would like our IE system to extract from "Anamnesis" (the selected part) information about events in "past tense" and "present continuous tense". Let us consider the selected text. As a result of the Annotation analysis and Chunking module, after the morphological and partial syntax analysis, we obtain for the given text the annotation on fig. 4.

In the paragraph "Захарният диабет е установен през 1998 год. на фона на наднормено тегло. Първоначално приемала Новонорм в комбинация със Сиофор, после Диапрел МР със Сиофор, но поради липса на ефект и изразени странични реакции към Метформин, от януари 2005 г. преминала на Инсулин Ново Микс в двукратни приложения." the IE system can recognise the following "keywords" signaling past events and event sequences - for tense (1998 год./year 1998, първоначално/at first , после/afterwards, януари 2005 г./january 2005), for effect - diagnosis


```

( (VPA (VPA (VPS (N (Ncmsf Захарният)) (Adv (Dt диабет)) (VPC (V (V (Vxif e))
(Participle (Vpptcv установен))) (PP (Prep (R през)) (N (Ncmpr 1998r)))))) (pt .))
(PP (Prep (R на)) (NPA (N (Ncmsh фона)) (PP (Prep (R на))
(NPA (A наднормено) (N (Ncmpr тегло)))))) )

( (CoordP (ConjArg (VPA (Adv (Dt Първоначално)) (VPC (V (Vpitf приемала))
(H (name (PNE Новонорм))) (PP (Prep (R в)) (NPA (NPA (NPA (N (Ncfsd комбинация))
(PP (Prep (R със)) (H (name (PNE Сиофор)))))) (NPA (pt .) (Adv (Dt после))
(H (name (NEPers (PNE Диапрел) (PNE MR)))))) (PP (Prep (R със))
(NPA (abbr Сиофор)))))) (Conj (pt .) (C (Cс но))) (ConjArg (VPA (PP (Prep (R поради))
(NPA (N (Ncfsi липса)) (PP (Prep (R на)) (CoordP (ConjArg (N (Ncmsi ефект))
(Conj (C (Cp и)) (ConjArg (NPA (NPA (A (Afsd изразени)) (NPA (A (Apsi странични))
(N (Ncfr реакци)) (PP (Prep (R към)) (NPA (N (NEOther (PNE Метформин))
(PP (Prep (R от)) (N (Ncmsi януари)))))))))) (VPC (V (Vpitf 2005r.)
(NPA (NPA (N (Ncfsd преминала)) (PP (Prep (R на))
(N (name (NEOrg (PNE Инсулин) (PNE Ново) (PNE Микс)))))) (PP (Prep (R в))
(NPA (A двукратни) (N (Ncmpr апликации)))))) )

```

Fig. 4. Text annotation after morphological and partial syntax analysis

(установен/determined), treatment (приемала/took, преминала на/changed to), symptoms (реакции/reactions, ефект/effect, наднормено/over).

As characteristics for these events, the system identifies drug entities after checking the medical goods list in the repository bank. The first event has no associated start time marker and the IE module is looking for time information in previous sentences in the text, by processing the local discourse. In case of lack of a marker the module takes into account the so called narrative convention - two past events, expressed consecutively have happened in the same order they are written in. Thus we obtain the filled templates shown in fig. 5. An important project task at present is to evaluate the accuracy of the partial temporal analysis which is sketched in this section.

6 Cause-Effect and Complex relations

On the next level of our knowledge representation model we aim at cause-effect relations extraction. As in the previous level several "keywords" for detection of such relations are used as well as manually prepared rules and templates for filling the appropriate information from the EHRs. These templates have slots like: Cause, Effect, Type, Degree, Evidence, Condition. The IE approach we apply is mostly based on pattern matching method. The keywords are classified into four main types based on the comprehensive typology of causal links of Altenburg [20] as follows: the *adverbial link* (e.g. hence, therefore), the *prepositional link* (e.g. because of, on account of), *subordination* (e.g. because, as, since, for, so) and *the clause-integrated line* (e.g. that is why, the result was). Causative verbs are transitive action verbs that express a causal relation between the subject and object or prepositional phrase of the verb. Some of these causal links were reconsidered for Bulgarian due to the specific language syntax and semantics.

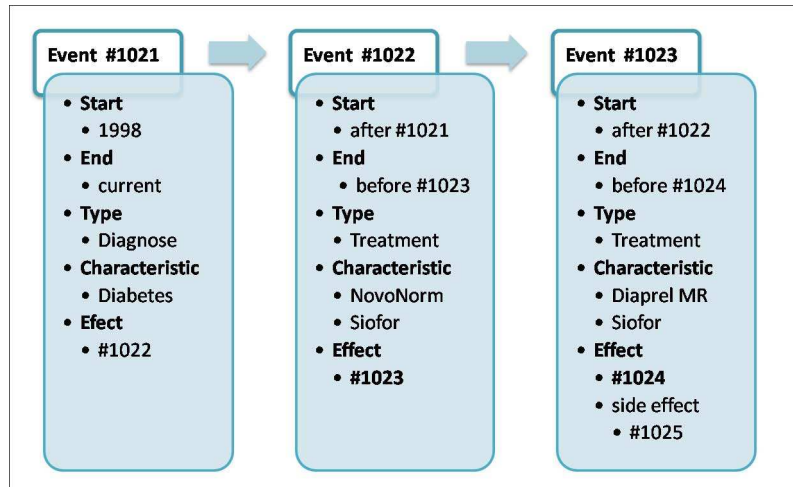


Fig. 5. Filled Temporal Conceptual Templates: Three consecutive events and their characteristics

The system has to be able to recognise several causal expressions (paraphrases) representing one and same causal event/situation.

On top of all structures recognised so far, the system infers implicit knowledge. Complex relations extraction is a challenging task that relates to recognition of implicit relations. This approach will help us to synthesise new knowledge by chaining extracted cause-effect relations. Here we make use of logical representation forms of the cause-effect relations as well as inference rules. At this level the processing is going beyond the single patient data and spreads to cause-effect relations, extracted from several patients' EHRs. These complex relations are statistically observed and additional relations are recognised. The same relations are later set as hypothesis which can be used in the cascade process as a basis for the next step of the inference procedure. Since there is no predefined expected result and the inference on all possible combinations of results is too complex and takes a long time for computation, it is necessary to set in advance the depth level of inference and type of the expected relations. It is also possible that the system in the unsupervised process extracts inappropriate results and that is why the inferred relations (hypothesis) need some level of supervised revision before they are saved in the system and further utilised.

7 Conclusion and Future Work

In this initial stage of the system design and implementation we discuss the obstacles in dealing with unstructured EHR texts and ideas for overcoming them by using language dependent techniques like related words, morphological analysis,

stemming. We propose bottom-up and top-down approaches for patient status observation and give directions for future development of the algorithms.

We discuss an algorithm for extracting temporal conceptual structures from the raw text corresponding to sequences of consecutive events representing the development of the patient disease. In this early stage we cannot discuss precision of our modules yet, but give outline for future development. Precise evaluation is a goal of one of the next project phases.

The project objective is to develop algorithms for discovering more complex cause-effect relations and other dependencies that are not explicitly given in the text. The modules for further analysis of more complex relations will be developed in the future project stages.

8 Acknowledgements

This work is a part of the project EVTIMA (Effective search of conceptual information with applications in medical informatics”) which is funded by the Bulgarian National Science Fund by grant No DO 02-292/December 2008.

The patient records for the project are kindly provided by the University Specialised Hospital for Active Treatment of Endocrinology Acad. I. Penchev, which is part of the Medical University Sofia.

References

1. Cunnigham, H., Information extraction - an user guide. Research Memo CS-99-07, Computer Science Department, University of Sheffield, 1999 (<http://www.dcs.shef.ac.uk/~hamish/IE>).
2. Rao, R.B., L.V. Lita, C. D. Raileanu, R. S. Niculescu. Medical Entity Extraction From Patient Data.see patent information at <http://www.faqs.org/patents/app/20080228769>.
3. Goldacre, M. L. Kurina, D. Yeates, V. Seagroatt and L. Gill. Use of large medical databases to study associations between diseases, Q J Med, 93 (10): 669-675, 2000.
4. Baud R. H. A natural language based search engine for ICD10 diagnosis encoding, Med Arh. 2004; 58 (1 Suppl 2): 79-80.
5. MedLEE - A Medical Language Extraction and Encoding System, <http://lucid.cpmc.columbia.edu/medlee/>
6. Lee, C.H., Khoo, C., and J.C. Na, Automatic identification of treatment relations for medical ontology learning: An exploratory study. In I.C. McIlwaine (Ed.), Knowledge Organization and the Global Information Society: Proc. of the Eighth International ISKO Conference. Germany: Ergon Verlag, 2004, pp. 245-250.
7. Christopher S. G. Khoo, Syin Chan, Yun Niu, Extracting Causal Knowledge from a Medical Database Using Graphical Patterns. In Proc. of ACL, Hong Kong, 2000.
8. Leroy, G., Chen, H., and Martinez, J.D., A Shallow Parser Based on Closed-class Words to Capture Relations in Biomedical Text. Journal of Biomedical Informatics (JBI) vol. 36, pp 145-158, June 2003.
9. Natural Language Processing in Medical Coding. White paper of Language and Computing (www.landcglobal.com). April 2004.

10. Boytcheva, S., A. Strupchanska, E. Paskaleva, and D. Tcharaktchiev, Some Aspects of Negation Processing in Electronic Health Records. In Proc. of International Workshop Language and Speech Infrastructure for Information Access in the Balkan Countries, 2005, Borovets, Bulgaria, pp. 1-8.
11. Rassinoux A.-M., R.H. Baud, J.-R. Scherrer. A multilingual analyser for medical texts. In: Tepfenhart, W.M., J. R. Dick and J. Sowa (Eds.) *Conceptual Structures: Current Practices*, Proceedings of the 2nd Int. Conf. on Conceptual Structures, Springer, LNCS Volume 835,1994, 84-96.
12. Votruba, P., S. Miksch, and R. Kosara, Linking Clinical Guidelines with Formal Representations. In Proc. - 9th Conf. on Artificial Intelligence in Medicine in Europe (AIME 2003), p. 152-157, Springer, 2003.
13. Croitoru M., Bo Hu, Srinandan Dashmapatra, Paul Lewis, David Dupplaw, Liang Xiao. A Conceptual Graph Description of Medical Data for Brain Tumour Classification, 15th International Conference on Conceptual Structures (ICCS 2007), Sheffield, UK.
14. Delugach, Harry S., An Approach To Conceptual Feedback In Multiple Viewed Software Requirements Modeling, Proc. Viewpoints 96: Intl. Workshop on Multiple Perspectives in Software Development, Oct. 14-15, 1996, San Francisco.
15. Smith B. J. , Harry S. Delugach: A Framework for Analyzing and Testing Requirements with Actors in Conceptual Graphs. ICCS 2006: 401-412
16. Pfeiffer, H.D. and R.T. Hartley: Semantic Additions to Conceptual Programming, in Proceedings of the Fourth Annual Workshop on Conceptual Graphs, Detroit, MI, 6.0.7-1 - 8 (1989).
17. Soderland, S., D. Aronow, D. Fisher, J. Aseltine, and W. Lehnert, Machine Learning of Text Analysis Rules for Clinical Records. CIIR Technical Report, University of Massachusetts Amherst, 1995.
18. Gaizauskas, R., M. Hepple, N. Davis, Y. Guo, H. Harkema, A. Roberts, and I. Roberts, AMBIT: Acquiring medical and biological information from text. In S. J. Cox, editor, Proceedings of the UK e-Science All Hands Meeting, UK, 2003.
19. Mani I., Recent Developments in Temporal Information Extraction, In Proceedings of RANLP-03, 2003.
20. Altenberg, B.. Causal linking in spoken and written English. *Studia Linguistica*, 1984, 38(1), 20-69.
21. Nakov, P., BulStem: Design and Evaluation of Inflectional Stemmer for Bulgarian. Proceedings of Workshop on Balkan Language Resources and Tools (1st Balkan Conference in Informatics). Thessaloniki, Greece,2003.