From Individual EHR Maintenance to Generalised Findings: Experiments for Application of NLP to Patient-Related Texts

Galia Angelova[1], Dimitar Tcharaktchiev[2], Svetla Boytcheva[1,3], Ivelina Nikolova[1],
Hristo Dimitrov[2], and Zhivko Angelov[1]

[1] Institute of Information and Communication Technologies,
Bulgarian Academy of Sciences, Sofia, Bulgaria
[2] University Specialised Hospital for Active Treatment of Endocrinology "Acad. I. Penchev"
(USHATE), Medical University – Sofia, Bulgaria
[3] American University in Bulgaria, Blagoevgrad, Bulgaria
{galia,iva}@lml.bas.bg,
{dimitardt,svetla.boytcheva}@gmail.com, h_dimitrov@yahoo.com,
angelov@adiss-bg.com

Abstract. Experiments in automatic analysis of free texts in Bulgarian hospital discharge letters are presented. Natural Language Processing (NLP) has been applied to medical texts since decades but high-quality results have been demonstrated only recently. The progress in automatic text analysis opens new directions for secondary use of Electronic Health Records (EHR). It enables also the design and development of software systems which provide better patient access to his/her health records as well as better maintenance of large EHR archives. We report about successful extraction of important patient-related entities from hospital EHR texts and consider several scenarios for application of NLP modules in healthcare software systems.

Keywords: Information Extraction, AI in Health Informatics.

1 Introduction

The performance of automatic text analysis gradually improves during the last decades but these systems are still rarely used outside the research groups where they have been developed [1]. There are several reasons for the slow penetration of the NLP technologies into practical settings: *(i)* developing high-quality NLP is very expensive as it requires much effort for collection of relevant language resources as well as for the design and implementation of software that might process various formats and styles of texts. Moreover, the medical domain is very large and the initial investments for the development of convincing NLP demonstrators are huge; *(ii)* the exploitation of NLP modules requires constant effort for supporting lexicon updates and tuning the systems to new linguistic constructions and text types; *(iii)* it is well

R. Kountchev & B. Iantovics (Eds.): *Adv. in Intell. Anal. of Med. Data & Decis.*, SCI 473, pp. 203–212.
DOI: 10.1007/978-3-319-00029-9_18

known that the technology works with high accuracy but not 100%; therefore some results might be erroneous and misleading.

On the other hand, the quick adoption of EHR worldwide implies constant growth of electronic narratives discussing patient-related information. Actually the most important findings about the patients are kept as free texts in various documents and languages. These text descriptions are oriented to human readers but the computers might process them for statistical analysis, research purposes, obtaining data to support decision making and health management etc. In this way the so called Information Extraction (IE) becomes the dominating paradigm of NLP application to biomedical texts. The main IE idea is to extract automatically important entities from free texts, with accuracy as high as possible, and to build software systems operating on these entities (skipping the non-analysed text fragments). NLP in general is viewed as a rather complex Artificial Intelligence task so IE is proposed as a technology at the middle between keyword search and deep text analysis; it focuses on surface linguistic phenomena that can be recognised without deep inference. It is expected that IE progress would enable radical improvements in the clinical decision support, biomedical research and the healthcare in general [2].

This chapter is structured as follows. Section 2 briefly reviews related work and presents state-of-the-art figures about IE accuracy for various medical entities. Section 3 discusses the particularities of Bulgarian hospital discharge letters and overviews the IE components that have been developed during the last 3 years. Section 4 sketches potential applications of the IE prototypes and their extensions. Section 5 contains the conclusion.

2 Related Work

We consider quite briefly some state-of-the-art IE results for English clinical texts. The IE performance is measured by *precision* (the percentage of correctly extracted entities among all recognized entities in the test set), *recall* (the percentage of correctly extracted entities among all available entities in the test set) and the harmonic mean *f-score*=2**Precision***Recall*/(*Precision*+*Recall*). Below we refer to these performance indicators in order to present the background results and the context where we develop our IE prototypes for Bulgarian language.

Recent systems for extraction of drug treatment achieve accuracy higher than 90%: f-score: for instance, 91.40% for drug name and 94.91% for dosage [3]; or f-score 89.9% for drug name and 93.6% for dosage [4]. The system MedEx extracts medication events with f-score 93.2% for drug names, 94.5% for dosage, 93.9% for route, and 96% for frequency [5].

The automatic assignment of ICD codes to diagnoses achieves 89.08% accuracy [6]. The three top systems in the coding competition presented in [6] processed the negation, hypernyms and synonyms in some way and exploited the semantic network of UMLS [7].

Patient status data are also recognized with high precision and recall. This is due to the fact that the IE systems can be carefully trained to identify relatively small sets of predefined words. For instance, the patient smoking status is classified into 5 categories by selecting sentences which contain the relevant information with f-score 92.64% [8].

Research on temporal IE from clinical texts is a relatively recent activity in medical informatics [9]. Temporal IE systems perform reasoning on temporal clinical data for therapeutic assessments; summarize data from temporal clinical databases, and model uncertainty in clinical knowledge and data [10]. Having in mind the complexity of the temporal IE tasks, some leading groups in biomedical NLP develop annotation schemes in order to unify the efforts for creation of training corpora which are tagged with temporal markers [11].

3 IE from Bulgarian Hospital Discharge Letters

Our experiments were performed on 6,200 discharge letters of patients with endocrine and metabolic diseases. The patient records have been anonymised in the USHATE hospital when exporting them to files for research purposes [12].

3.1 Material

Discharge letters in Bulgarian hospitals traditionally contain predefined sections due to centralised regulations which date back to the middle of the last century. Sometimes the sections might be merged, skipped (when they are empty), their headers and the default section sequence might be changed. However, practically the most important sections are always included in the discharge letters: the sections *Diagnoses*, *Anamnesis*, *Patient Status*, *Lab data & clinical tests* and *Debate* are available in 100% of the letters in a training corpus of 1,300 EHR; the section *Past diseases* is available in 88.52% of the letters in the same corpus while the *Family Medical History* and *Allergies & risk factors* are included correspondingly in 52,22% and 43,56% of the EHRs. Using a preliminary prepared list of section headers (about 80 keywords and phrases), most available sections can be automatically recognised with accuracy 99,99%. This enables a precise splitting of the record texts into subsections. We note that structured discharge letters are relatively rare in the healthcare practice worldwide as this requires a long and stable centralized administrative tradition in the preparation of medical documentation.

Another interesting fact concerns the wordforms used in the Bulgarian discharge letters. Table 1 shows some statistics about occurrences of Bulgarian and Latin words and terms. It turns out that 37% of all words in the discharge letters are 'unknown' as they are Latin. About one half of the Bulgarian words are general lexica while many specific medical terms are not available in the usual (electronic) dictionaries. These figures illustrate the claims that high-quality NLP requires much effort for the development of extensive linguistic resources. Our only electronic resource with Bulgarian medical terms is the International Classification of Diseases (ICD-10) which contains 10970 terms.

Table 1. Distribution of words in the texts of 6,200 discharge letters

Terms and words	Wordforms	Basic words	Basic words / total
Bulgarian terms	601233	12009	63%
Latin with Latin alphabet	18926	560	3%
Latin terms transliterated to Cyrillic alphabet	179589	6465	34%
Total	**799748**	**19034**	100%

The free texts of patient records in our training corpus contain telegraphic phrasal descriptions and incomplete sentences, which excludes the application of deeper syntactic analysis and traditional NLP techniques in general. Therefore, in our IE experiments, we extract some important patient-related entities only and integrate them into more complex scenarios for decision making and structuring patient history.

3.2 Methods

We have developed several extracting modules for various clinical entities in two research projects. Figure 1 illustrates the types of patient-related entities that are extracted at present: diagnoses, drugs, entities from the Anamnesis section, patient status attributes as well as values of clinical test and lab data.

In the project PSIP (Patient Safety through Intelligent Procedures in medication, 2010-2011, PSIP FP7 ICT eHealth: http://www.psip-project.eu) our team extracted *(i)* drugs, *(ii)* diagnoses and *(iii)* values of clinical tests from USHATE hospital EHRs in order to fill in a PSIP-compliant repository and to validate the PSIP rules for Adverse Drug Events (see the bottom interface (d) on Fig 1). Most data items were available in the Hospital Information System (HIS) of USHATE and these items were sent to the PSIP repository directly by the HIS. But USHATE is a specialised hospital which treats endocrine diseases and their complications; thus drugs for accompanying and chronic diseases are often brought in by the patient and taken without records in the HIS. In this way it turned out that a considerable amount of the necessary data is presented only in the free text of the hospital discharge letters. Therefore we have implemented the following IE prototypes which process discharge letters that are split into sections [13]:

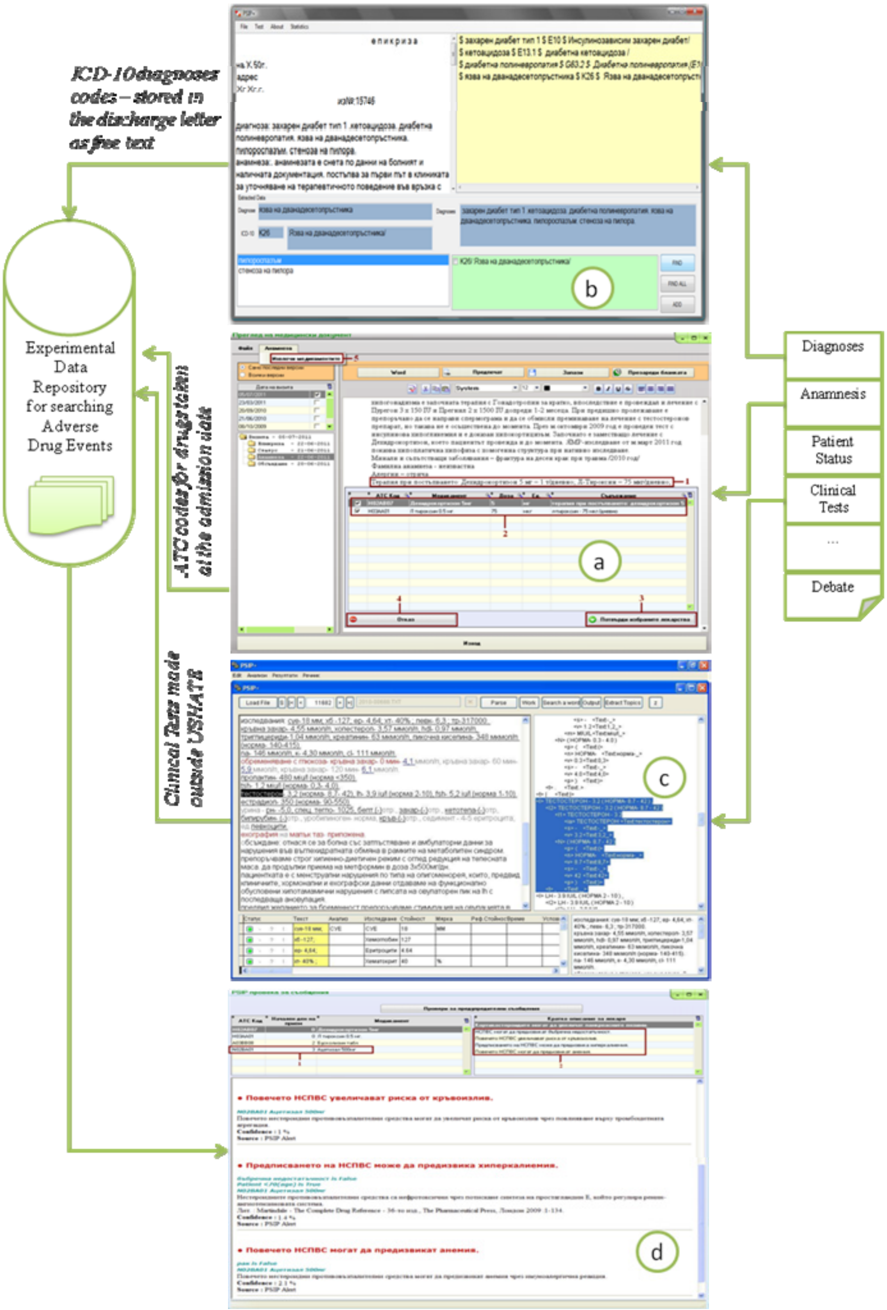

ICD-10 diagnoses codes – stored in the discharge letter as free text
Experimental Data Repository for searching Adverse Drug Events
ATC codes for drugs taken at the admission date
Clinical Tests made outside USHATE
Diagnoses
Anamnesis
Patient Status
Clinical Tests
...
Debate
a
b
c
d

Fig. 1. Extraction of important patient-related entities for supporting clinical decision making and structuring patient histories

(a) Extractor of treatment events, based on ATC[1] codes: the extraction works in several steps, starting by preprocessing which identifies phrases containing drug names, dosages, mode of admission and frequency within their contexts. Some 160,892 occurrences of drug names were automatically recognised in the texts of 6,200 EHRs. The extractor assigns the corresponding ATC code to each medication event. It takes into account negative statements, some elliptical constructions, typical conjunctive phrases, and makes simple inferences concerning temporal constraints. The extraction accuracy (f-score) for drug names is 98.42% and for dose - 93.85% [14]. In general the EHR texts might discuss past, present and future medication events but it is important to allocate the treatment which is relevant for the hospitalisation period. The extractor achieved f-score of 90,17% for the recognition in the *Anamnesis* texts of 355 drugs, taken by the patients during the period of hospitalisation, which are not prescribed via the Hospital Pharmacy [15]. This extractor was online integrated in the HIS during the PSIP validation in USHATE, see interface (a) on Fig. 1.

(b) Extractor for assignment of ICD-10 codes of diagnoses. Bulgarian hospitals are reimbursed by the National Insurance Fund via the 'clinical pathways' scheme. When a patient is hospitalised, medical experts might select from the HIS menu one diagnosis which is sufficient for the association of the desired clinical pathway to the respective patient. Thus most diseases are not formally diagnosed and they are entered in the EHR section *Diagnoses* as free text. We have developed an extractor which matches the phrases, listed in the *Diagnoses* section, to disease names as given in ICD-10 nomenclature [16], see interface (b) on Fig. 1. This extractor processes Bulgarian and Latin terms including Latin terms transliterated to Cyrillic. The major challenge in automatic encoding is that the correspondence between the EHR diagnoses and ICD-10 names/codes is not a one to one correspondence. There is a variety of linguistic expressions which might refer to the same diagnosis and one expression might refer to many diagnoses. For 6,200 EHRs, some 26,826 phrases were found in the *Diagnoses* sections; the extractor assigns correctly ICD-10 codes to 84,5% of them.

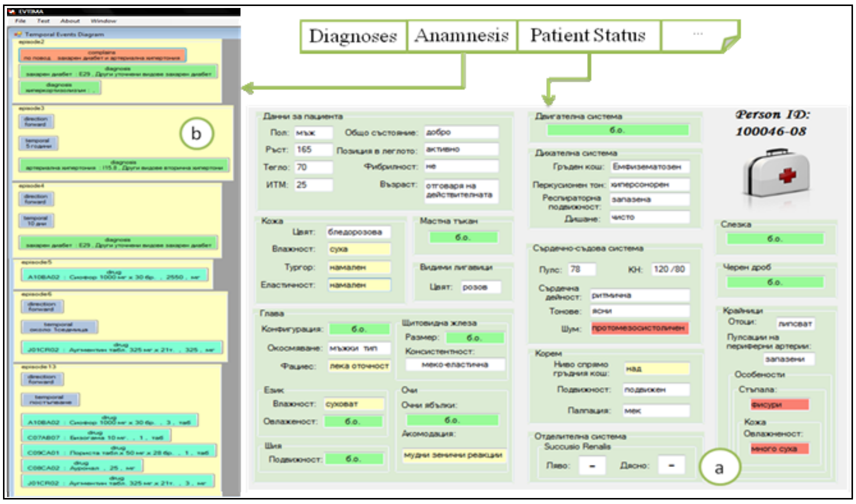
(c) Extractor of values of clinical tests and lab data. When a patient is examined in USHATE, the lab data are entered to the EHR via the HIS. However there are a number of tests made outside the hospital and their results need to be extracted from the free text of *Lab data & clinical tests* sections (see interface (c) on Fig. 1.). This extractor works with precision of 98,2%.

The research project EVTIMA[2] (Efficient Search of Conceptual Patterns with Application in Medical Informatics) deals with IE and structuring of patient descriptions in order to build an internal conceptual representation which enables quick search of repeating patterns in diabetes development. Initially we considered the status descriptions as they are narrated in the *Patient Status* sections [17]. Skin status is extracted with f-score 83.33%, neck status – with f-score 91.67%, thyroid gland status – with f-score 92.59%, and limb status – with f-score 89.01%. Age descriptions are extracted

[1] Anatomical Therapeutic Chemical (ATC) Classification System, `http://atc.thedrugsinfo.com/`

[2] Funded by the National Science Fund in Jan. 2009-June 2012, see `http://www.lml.bas.bg/evtima`

from anonymised EHRs with f-score 89.44%. Status extraction needs explicitly-declared domain knowledge which supports the IE algorithms by providing constraints and inference mechanisms. A panel capturing the structured patient status is shown on Fig. 2 interface (a) on the top.



episode2
complains: по повод захарен диабет и артериална хипертония
diagnosis: захарен диабет : E29 . Други уточнени видове захарен диабет
diagnosis: хиперхолестеролемия
episode3
direction: forward
temporal: 5 години
diagnosis: артериална хипертония : I15.8 . Други видове вторична хипертония
episode4
direction: forward
temporal: 10 дни
diagnosis: захарен диабет : E29 . Други уточнени видове захарен диабет
episode5
drug: A10BA02 : Сиофор 1000 мг x 30 бр. . 2550 . мг
episode6
direction: forward
temporal: около 1седмица
drug: J01CR02 : Аугментин табл. 325 мг x 21т. . 325 . мг
episode13
direction: forward
temporal: постъпване
drug: A10BA02 : Сиофор 1000 мг x 30 бр. . 3 . таб
drug: C07AB07 : Бисогамма 10 мг . . 1 . таб
drug: C09CA01 : Лориста табл.x 50 мг x 28 бр. . 1 . таб
drug: C08CA02 : Агрофан . 25 . мг
drug: J01CR02 : Аугментин табл. 325 мг x 21т. . 3 . мг

Данни за пациента
Пол: мъж
Ръст: 165
Тегло: 70
ИТМ: 25
Общо състояние: добро
Позиция в леглото: активно
Фибрилност: не
Възраст: отговаря на действителната

Кожа
Цвят: бледорозова
Влажност: суха
Тургор: намален
Еластичност: намален

Мастна тъкан
б.о.

Видими лигавици
Цвят: розов

Глава
Конфигурация: б.о.
Окосмяване: мъжки тип
Фациес: лека оточност

Щитовидна жлеза
Размер: б.о.
Консистентност: меко-еластична

Език
Влажност: суховат
Овлаженост: б.о.

Очи
Очни ябълки: б.о.
Акомодация: мудни зенични реакции

Шия
Подвижност: б.о.

Двигателна система
б.о.

Дихателна система
Гръден кош: Емфизематозен
Перкусионен тон: хиперсонорен
Респираторна подвижност: запазена
Дишане: чисто

Сърдечно-съдова система
Пулс: 78
КН: 120 /80
Сърдечна дейност: ритмична
Тонове: ясни
Шум: протомезосистоличен

Корем
Ниво спрямо гръдния кош: над
Подвижност: подвижен
Палпация: мек

Отделителна система
Succusio Renalis
Ляво: -
Дясно: -

Person ID: 100046-08

Слезка
б.о.

Черен дроб
б.о.

Крайници
Отоци: липсват
Пулсации на периферни артерии: запазени
Особености
Стъпала: фисури
Кожа
Овлаженост: много суха

Fig. 2. Extraction of structured patient history

Our recent research is focused on automatic construction of temporal event sequences from the *Anamnesis* sections. We have developed a conceptual model of episodes that happen or occur to the patient; the description of these episodes is summarised in the *Anamnesis* section. When a diabetic patient is hospitallised, the medical experts briefly document the timeline of diabetes progression and the major events in the patient history. Starting with the moment of assigning the principal diagnose and following the complications, they summarise the drugs and procedures that were applied along the years. Episodes have *beginning* and *end* (sometimes without clearly defined time moments) and can be recognised after identification of about 80 types of temporal markers (prepositional and adverbial phrases). About 83.36% of the temporal markers refer to explicitly specified moments of time and can be seen as *absolute* references [18]. The remaining markers express *relative* or *undetermined* references. The markers are identified with precision 87% and recall 68%. The direction of time for the episode events: backwards or forward (with respect to certain moment orienting the episode) is recognised with precision 74.4%. The events happening within the episodes include *(i)* drug admission, *(ii)* diagnosing a disease or complication and *(iii)* changes of the patient condition or status. The extractors for drugs and diagnoses are integrated into the episode structuring. Identifying diabetes symptoms and conditions in free text is uneasy task as no 'canonical forms' exist in any dictionary for the related phrases and paraphrases. Therefore the currently developed extractor achieves lower precision. For instance, blood sugar level and body weight change are identified with f-score 60-96% in the separate processing phases [19]. Fig. 2 (b) on the left displays temporal splitting of patient history into episodes.

4 Potential Applications of the IE Prototypes

The results presented in section 3 were obtained in research projects but they explicate some specific particularities of the contemporary medico-administrative practice in Bulgaria. Having analysed the texts of only 6,200 discharge letters, we see for instance that:

(i) *Diagnoses are not fully encoded in the HIS*. There are 9,321 diagnoses registered in the HIS for the 6,200 EHRs in the test corpus while our extractor found in the texts 22,667 phrases to which ICD-10 codes were assigned;

(ii) *Drugs are not fully encoded in the hospital pharmacy.* According to the HIS entries, there are in average 1,9 drugs per patient, while our drug extractor found in average 5,9 drug names related to the period of hospitalization. The patients in USHATE took some 355 drugs which they brought to the hospital privately (while the hospital pharmacy of USHATE operated with 1182 drugs during the period of our experiments);

(iii) *Many clinical tests are done outside the hospital*, e.g. the hormonal ones, and their results are manually retyped in the corresponding discharge letters; thus the test results can be analysed only by human readers.

These examples show that the NLP extractors, when run over statistically-significant data excerpts, would enable important observations that are impossible at present; for instance the diagnose and drug extractors would enable to quickly construct diagnostically-related homogeneous groups for patients who need to be treated in a similar way. Definition of such groups would facilitate the optimisation of the drug treatment and reimbursement of costs. In this way NLP supports secondary use of the EHR data since it enables quick production of large data resources that might be extracted from millions of patient-related texts.

Regarding the potentially erroneous NLP results, we note that human data processing is not perfect as well. For instance the manual encoding of ICD-10 diagnoses might be erroneous too; [20] reports about 48-49% errors of human coders during the first year of their practice which decrease to 7% errors when the coding experts gain experience. These figures show that the results of our extractor, which assigns ICD-10 codes to diagnoses with accuracy of 84.5%, are within the margins of the usual correctness that is achieved in complex medical tasks as diagnose encoding.

Another direction of potential NLP application is to support the patients when they access their EHR data. The European Commission recommends that the patients should be able to inspect (part of) the content of their EHRs. Having in mind the Bulgarian tradition to use Latin terms both in Latin and Cyrillic alphabets, the first possible application is to ensure automatic translation of the diagnose terms to Bulgarian disease names (we have shown that there are 37% Latin terms in the EHR texts). In this way NLP might provide the 'normalisation' of the text and facilitate the active participation of the patients in their treatment. Last but not least NLP might support also the patient inclusion in the monitoring of his/her treatment by generation of alerts, explanations and recommendations.

5 Conclusion

The technologies for automatic text analysis are developing and deliver gradually improving results which encourage plans for practical NLP application. Despite the fact that much effort is needed to attach complex linguistic phenomena like temporality, negation, conditionality, reference resolution as well as semantic and pragmatic interpretation, we find running projects for biomedical text processing in all countries with advanced information societies. Basic sets of medical terminology are supported in hundreds of natural languages by the World Health Organisation which facilitates the development of electronic dictionaries suitable for NLP. The growing amount of clinical narratives is another stimulating factor that implies the increasing interest in automatic text processing.

Our experience shows that the IE technology is sufficiently mature to be integrated in practical software applications as a tool for delivery of large amounts of extracted entities. The application niches need to be carefully selected and the IE tools should be properly integrated. We are focused on diabetes as a major chronic disease consuming significant budget by the Health Insurance Fund. Our present results show the potential of IE to provide data to healthcare managers and decision makers concerning optimisation of diabetes treatment and reimbursement. Current research was set for Bulgarian Language, but similar methods were successfully used also for other languages in PSIP project.

Acknowledgments. The research tasks leading to these results have received funding from the EC's FP7 ICT under grant agreement n° 216130 PSIP (Patient Safety through Intelligent Procedures in Medication) as well as from the Bulgarian National Science Fund under grant agreement n° DO 02-292 EVTIMA (Efficient Search of Conceptual Patters with Application in Medical Informatics).

References

1. Meystre, S., Savova, G., Kipper-Schuler, K., Hurdle, J.F.: Extracting Information from Textual Documents in the EHR: A Review of Recent Research. In: Geissbuhler, A., Kulikowski, C. (eds.) IMIA Yearbook of Medical Informatics, pp. 138–154 (2008)
2. Demner-Fushman, D., Chapman, W., McDonald, C.: What can NLP do for Clinical Decision Support? J. of Biomedical Informatics 42(5), 760–772 (2009)
3. Patrick, J., Li, M.: A Cascade Approach to Extracting Medication Events. In: Proc. Australian Language Technology Workshop (ALTA), pp. 99–103 (2009)
4. Halgrim, S., Xia, F., Solti, I., Cadag, E., Uzuner, Ö.: Extracting Medication Information from Discharge Summaries. In: Louhi 2010, Proc. of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents, pp. 61–67 (2010)
5. Xu, H., Stenner, S., Doan, S., Johnson, K., Waitman, L., Denny, J.: MedEx: a Medication Information Extraction System for Clinical Narratives. J. Am. Med. Informatics Assoc. (17), 19–24 (2010)

6. Pestian, J., Brew, C., Matykiewicz, P., Hovermale, D.J., Johnson, N., Cohen, K.B., et al.: A Shared Task Involving Multi-label Classification of Clinical Free Text. In: ACL 2007 Workshop on Biological, Translational, and Clinical Language Processing (BioNLP 2007), Prague, pp. 36–40 (2007)
7. UMLS, the Unified Medical Language System, `http://www.nlm.nih.gov/research/umls`
8. Savova, G., Ogren, P., Duffy, P., Buntrock, J., Chute, C.: Mayo Clinic NLP System for Patient Smoking Status Identification. J. Am. Med. Inform. Assoc. 15, 25–28 (2008)
9. Zhou, L., Hripcsak, G.: Temporal Reasoning with Medical Data - a Review with Emphasis on Medical NLP. J. Biom. Informatics 40(2), 183–202 (2007)
10. Adlassnig, K.-P., Combi, C., Das, A., Keravnou, E., Pozzi, G.: Temporal Representation and Reasoning in Medicine: Research Directions and Challenges. AI in Medicine 38(2), 101–113 (2006)
11. Savova, G., Bethard, S., Styler, W., Martin, J., Palmer, M., Masanz, J., Ward, W.: Towards Temporal Relation Discovery from the Clinical Narrative. In: Proc. AMIA Annual Symposium 2009, pp. 568–572 (2009)
12. Nikolova, I., Dimitrov, H., Tcharaktchiev, D.: Ethics and Security in Text Mining of Patient Records in Bulgarian: the EVTIMA Solution. In: ACM Proceedings of CompSysTech (2010)
13. Tcharaktchiev, D., Angelova, G., Boytcheva, S., Angelov, Z., Zacharieva, S.: Completion of Structured Patient Descriptions by Semantic Mining. In: Koutkias, V., Niès, J., Jensen, S., Maglaveras, N., Beuscart, R. (eds.) Studies in Health Technology and Informatics, vol. 166, pp. 260–269. IOS Press (2011)
14. Boytcheva, S.: Shallow Medication Extraction from Hospital Patient Records. In: Koutkias, V., Nies, J., Jensen, S., Maglaveras, N., Beuscart, R. (eds.) Studies in Health Technology and Informatics, vol. 166, pp. 119–128. IOS Press (2011)
15. Boytcheva, S., Tcharaktchiev, D., Angelova, G.: Contextualization in Automatic Extraction of Drugs from Hospital Patient Records. In: Moen, A., et al. (eds.) Proc. of MIE-2011, the 23rd Int. Conf. of EFMI, Studies in Health Technology and Informatics, Norway, vol. 169, pp. 527–531. IOS Press (August 2011)
16. Boytcheva, S.: Automatic Matching of ICD-10 Codes to Diagnoses in Discharge Letters. In: Proc. of Biomedical NLP Workshop, Satellite Event of Int. Conf. RANLP 2011, pp. 19–26 (2011), `http://aclweb.org/anthology-new/W/W11/W11-42.pdf`
17. Boytcheva, S., Nikolova, I., Paskaleva, E., Angelova, G., Tcharaktchiev, D., Dimitrova, N.: Obtaining Status Descriptions via Automatic Analysis of Hospital Patient Records. Special Issue on Semantic IT of Informatica, Int. J. of Computing and Informatics (Slovenia) 34(4), 269–278 (2010); Fomichov, V. (ed.)
18. Boytcheva, S., Angelova, G., Nikolova, I.: Automatic Analysis of Patient History Episodes in Bulgarian Hospital Discharge Letters. In: Proc. Demonstrations at the EACL 2012, pp. 77–81. ACL, France (2012), `http://www.aclweb.org/anthology/E12-2016`
19. Nikolova, I.: Unified Extraction of Health Condition Descriptions. In: Proc. of the NAACL HLT 2012 Student Research Workshop, pp. 23–28. ACL, Montreal (2012), `http://www.aclweb.org/anthology/N12-2005`
20. Ivanov, L., Ganova-Yolovska, M., Konstantinov, B.: Quality of Coding and Reliability of Medical Information for Distribution of Financial Resources into Diagnostically-related Groups. Social Medicine 4, 32–34 (1999) (in Bulgarian)