Tool for Automatic Annotation of Clinical Texts in Bulgarian – BGMedAnno

Sylvia Vassileva [1], Svetla Boytcheva [2] and Ivan Koychev [1]

[1] *Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski", Sofia, Bulgaria*
[2] *Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria*

Abstract

This paper describes the design of BGMedAnno: an automatic annotation tool for clinical text in Bulgarian. The proposed solution combines classical rule-based and dictionary-based approaches for name entity recognition (NER) with more advanced deep learning methods to identify different categories of medical terms, as well as some nested objects. The following categories of medical terms are currently identified: symptoms, complaints, diagnoses, anatomical organs and systems, risk factors, and family history. In addition, the negation relation and its scope are recognized. The location relation was modeled for connecting different categories of symptoms and complaints to anatomical organs and systems. All identified concepts were normalized to medical standard classifications and ontologies like UMLS, MESH, ICD-10 and mapped into the concepts of the linked open data Knowledge Graph WikiData. The proposed approach for automatic medical terms and their relations recognition shows high accuracy. The rule-based method shows an F1-score of 75%, while the trained BERT-based model presents an F1 score of 73%. Although the BERT model performs slightly worse on the test set, observations show that it finds objects in sentences that are not covered by the rule-based method. For the object linking task, the developed method based on the BERT language model shows a 61% F1 result, significantly outperforming direct string comparison, which achieves an F1 result of 45%. The developed user interface allows direct application of the annotation tool for individual texts. The service API outputs data in JSON format, which enables interoperability with other systems and can be used to process large collections of clinical data.

Keywords

Annotation tools, natural language processing, health informatics, deep learning, machine learning

Information Systems & Grid Technologies: Fifteenth International Conference ISGT'2022, May 27–28, 2022, Sofia, Bulgaria
EMAIL: sylvia.vassileva@gmail.com (S. Vassileva); svetla.boytcheva@iict.bas.bg (S. Boytcheva); koychev@fmi.uni-sofia.bg (I. Koychev)
ORCID: 0000-0002-5542-9168 (S. Boytcheva); 0000-0003-3919-030X (I. Koychev)

1. Introduction

Medical term recognition and linking to corresponding entities in knowledge bases, standard classifications, and ontologies is a very important task in clinical text analysis. It allows extracting structured data from clinical text and the subsequent processing of the extracted information. Structured clinical data is useful in many real-life scenarios like automatic analysis of drug effects in clinical trials, extracting hidden relations between risk factors and diseases, searching for similar patient cases, as well as clinical decision support systems. By improving the clinical software systems, clinicians can focus and improve patient care.

In this paper, we present BGMedAnno[2]: an automatic annotation tool of clinical text in Bulgarian. We propose an innovative approach for named entity recognition (NER) of medical terms (symptoms, complaints, diagnoses, anatomical organs, and systems) and their automatic linking to relevant concepts from the Unified Medical Language System (UMLS[3]), Medical Subject Headings (MeSH[4]) and International Classification of Diseases (ICD-10-CM[5]).

2. Related work

There are many tools for manual annotation of clinical text [1], but automatic annotation continues to be a challenging task. The automatic annotation tools perform two main tasks: Named Entity Recognition (NER) and relations extractions (RE). Most systems divide the process into two separate steps, choosing a different approach for each of them. In contrast, others perform an end-to-end process that discovers and links concepts simultaneously. The classical NER systems were based on manually created specific rules and features, appropriate to a specific domain. The current state-of-the-art systems are based on deep neural networks using vector representations of text [2]. The main approaches to finding objects in a text can be organized into the following groups: rule-based, unsupervised Machine Learning (ML), supervised ML, and deep learning.

2.1. Rule-based and dictionary-based NER systems

The rules are manually created using dictionaries and templates, which is a time-consuming laborious complex task that requires domain-specific knowledge. This makes these approaches suitable mainly for some specific domains. The advantage of this approach is that it does not require annotated training data.

[2] https://bgmedanno.fmi.uni-sofia.bg

[3] https://www.nlm.nih.gov/research/umls/index.html

[4] https://www.ncbi.nlm.nih.gov/mesh

[5] https://www.cdc.gov/nchs/icd/icd10cm.htm

It works very well if the dictionaries are complete. In the biomedical domain, synonyms of classifications such as UMLS, MeSH, and ICD-10-CM can be used. For example, Savova et al [3] propose a dictionary search approach built from terms from UMLS, SNOMED CT[6] and RxNORM[7], showing an F1 score of 71.5% on Mayo Clinic EMR (cTAKES) data. Chen's system [4] is also rule-based, trained on n2c2-1 [5] competition data with an F1 score of 90%. In the clinical field in Bulgarian, Todorova [6] implements a system based on rules and dictionaries, which recognizes symptoms, complaints, organs, anatomical systems, and other categories and demonstrates an F1 score of 91.4% on a small dataset of samples of anonymized outpatient records (ORs) from the Bulgarian National Diabetes Registry [7].

2.2. NER using unsupervised ML

These approaches do not require manually annotated data. Very often these systems use clustering, which categorizes the objects in the text based on their proximity. Clusters are defined based on lexical resources, templates, and statistics from large corpora [2]. In the biomedical domain, Zhang et al [8] proposed an unsupervised approach using dictionaries of terms, corpus statistics such as IDF and contextual vectors, and shallow syntactic knowledge (using noun phrases). Experiments conducted with data from GENIA and i2b2-Pittsburgh show better results than other unsupervised approaches – 15.2% and 26.5% micro-F1 with a complete match, but worse results than the supervised approaches.

2.3. NER using supervised ML

In supervised ML, the NER task can be considered as a multiclass classification task of each word or as a sequence labeling task. Classical algorithms for supervised ML require careful selection of the input properties of the text. Each word in the text is represented as a vector, which may consist of one or many boolean, numerical, or nominal values. This group of approaches often uses properties of words such as morphology, parts of speech, as well as properties of the document or corpus as word frequency. The review of Li et al [2] gives examples of common algorithms in this approach – Hidden Markov Models (HMM), decision trees, decision entries, maximum entropy models, Support Vector Machine (SVM), and Conditional Random Fields (CRF). McNamee and Mayfield [9] consider the task as a classification of each word of the document in one of 8 classes and train one SVM classifier per class. This approach does not use the word context to make a decision. In contrast, the CRF approach views the task as a sequence labeling task and uses context. This approach is used in many articles for

[6] https://www.snomed.org

[7] https://www.nlm.nih.gov/research/umls/rxnorm/index.html

NER, including in the biomedical domain: McCallum and Li [10] used the CRF approach and showed an F1 score of 84.04% on CoNLL03 data; Settles [11] uses CRF for biomedical sites and shows an F1 of 69.5% on data from the BioNLP / NLPBA 2004 shared task. The main challenge with this approach is the need for annotated training data.

2.4. NER based on Deep Learning (DL)

With these approaches, deep neural networks automatically detect important text properties that help object identification. A major challenge in this approach is the need for a large amount of annotated training data. Many systems in the biomedical domain are based on DL approaches: Magge et al [12] use a bidirectional LSTM model with a CRF output layer, which achieves a macro F1 score of 81%; Wu et al [13] compared CRF models with CNN and RNN networks and showed that RNN performed best on i2b2 2010 data, achieving an F1 score of 85.94%; Vunikili et al [14] uses BERT [15] and Spanish BERT (BETO) [16] for transfer learning. The model derived tumor information from clinical records in Spanish and achieved an F1 score of 73.4%.

3. Clinical data and medical ontologies

3.1. Ontologies and standard medical classifications

3.1.1. Unified Medical Language System (UMLS)

The Unified Medical Language System (UMLS) is an international system of vocabularies, classifications, and standards for encoding medical terms. It aims to facilitate the creation and integration of biomedical information systems and services. UMLS consists of three parts – meta-thesaurus, semantic network, and specialized lexicons and tools. The meta-thesaurus consists of concepts and their corresponding encoding in different systems like CPT[8], ICD-10-CM, LOINC[9], MeSH, RxNORM, and SNOMED CT. It contains more than 5 million terms, related to concepts with unique identifiers. Each concept has a relationship with one or more lexical variations of the term. For example, the term “atrial fibrillation” has a code C4015486. Unfortunately, UMLS does not have an official translation in Bulgarian.

3.1.2. Medical Subject Headings (MeSH)

Medical Subject Headings (MeSH) is a thesaurus indexing PubMed[10] articles, consisting of more than 33 million biomedical articles from Medline, scien-

[8] https://www.ama-assn.org/practice-management/cpt

[9] https://loinc.org

[10] https://pubmed.ncbi.nlm.nih.gov

tific journals, and online books. MeSH contains about 27,000 terms, structured in a hierarchy. Each MeSH descriptor uniquely identifies a concept and can appear in one or more places in the hierarchy. The code for "atrial fibrillation" in MeSH is D001281. Unfortunately, MeSH does not have an official translation in Bulgarian.

3.1.3. International Classification of Diseases 10th revision

The International Classification of Diseases 10th revision is a statistical classification of diseases used worldwide. It is used in the Bulgarian health insurance domain for statistical and reimbursement purposes. It is a hierarchical classification grouping diseases in several different levels.

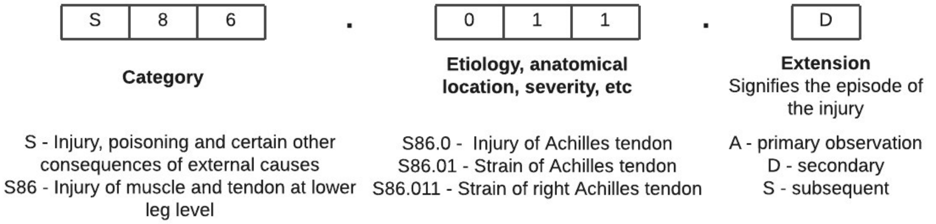

S
8
6
.
0
1
1
.
D
Category
Etiology, anatomical location, severity, etc
Extension
Signifies the episode of the injury
S - Injury, poisoning and certain other consequences of external causes
S86 - Injury of muscle and tendon at lower leg level
S86.0 - Injury of Achilles tendon
S86.01 - Strain of Achilles tendon
S86.011 - Strain of right Achilles tendon
A - primary observation
D - secondary
S - subsequent

Figure 1: Structure of ICD-10 codes

Figure 1 shows the structure of ICD-10 codes – they can be 3 to 7 characters in length depending on the specificity of the classification. ICD-10 codes of 4 characters are used in Bulgaria. The number of 4-character codes is almost 11,000. For this paper, we use the 4-character codes as they are most relevant for Bulgarian healthcare. Each disease can be represented with one or more codes, depending on the underlying cause or its localization. For example, "diabetic polyneuropathy, retinopathy, and neuropathy" can be encoded as G36.2 as a disease of the nervous system or H36.0 as a retinal disease. There is an official ICD-10 translation in Bulgarian which is used in this paper.

3.2. Clinical texts

3.2.1. Medical term vocabularies

For this paper, we gather medical term vocabularies in different categories using the following sources: vocabularies gathered by Todorova [6]; website of Alexandrovska hospital[11]; website Puls[12]; diagnosis dataset by Boytcheva et al [17] – more than 170,000 unique diagnoses and their ICD-10 codes.

[11] https://www.alexandrovska.com/display.php

[12] https://www.puls.bg/diagnostic/symptom

Table 1 shows the different term categories and the number of terms in each vocabulary. The total number of terms in all vocabularies is 180,553.

Table 1
Vocabulary categories and the number of terms per category

Category	Description	Number of terms
DIAGNOSIS	Diseases	177,271
SYMPTOM	A doctor's assessment of the patient's problems	762
COMPLAINT	A patient's subjective report of their problems	1,112
ORGAN	Human Organs	1,213
ANATOMICAL SYSTEM	Anatomical systems in the human body	38
FAMILY	Family relation types	53
FAMILY HISTORY	Family history phrases	27
RISK FACTOR	Risk factors to health	51
NEGATION	Negative phrases	26

3.2.2. Patient history dataset

We extract patient history records from anonymized patient data including hospital discharge letters and outpatient records. All names and dates were removed in advance and the patient history records were split into sentences and shuffled, resulting in 41,066 records. The average number of words in a record is 21.25 words. This dataset does not have any labels. To generate labels, we use the approach based on rules and dictionaries by Todorova [6].

As a result, we have a labeled corpus of patient history records. Since this method is automatic, there can be some entities that were missed or annotated incorrectly.

Figure 2 shows the number of labeled entities using the automatic approach.

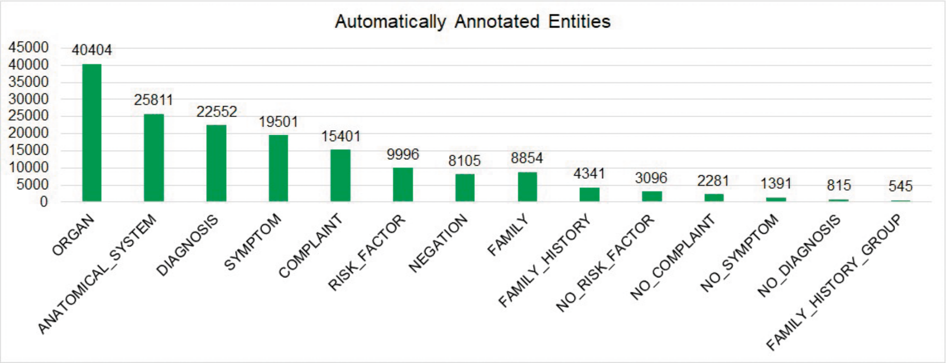


Figure 2: Automatically annotated entities per category

3.2.3. Medical term knowledge base

A knowledge base from WikiData was gathered including diseases, organs, anatomical systems, symptoms, and complaints which have UMLS, MeSH, or ICD-10 code. The knowledge base contains 122,922 terms related to 21,174 concepts. Since not all WikiData terms are translated in Bulgarian, we use Google Translate to translate the English terms automatically and then manually review and clean the translation. Some terms are part of more than one classification and/or can have one or more codes related to them. Table 2 shows the number of examples and concepts from each system:

Table 2
The number of examples and concepts from each medical system in the knowledge base

Medical System	Number of examples	Number of concepts
UMLS	114,883	21,481
MeSH	63,181	9,422
ICD-10	58,175	4,073

3.2.4. Outpatient records dataset

A small dataset of outpatient records was manually labeled with medical terms and their corresponding entities in WikiData. The outpatient records are written by General Practitioner physicians. The dataset consists of 30 records that are used for testing purposes. Nested entities are labeled in the following categories: organs, anatomical systems, diagnosis, symptoms, complaints, family history, risk factors, and negations. The data contains 170 sentences and 582 labeled entity tokens. The number of labeled concepts from WikiData is 446 words linked to 111 unique entities.

4. Methods for medical terms recognition and linking

We split the task for recognizing and linking medical terms into two parts – the first is to recognize the term mentioned in the text and assign a category (named entity recognition) and the second is to find the corresponding concept in the knowledge base (entity linking). The overall architecture of the tool BG-MedAnno is shown in Figure 3. Each of the three main tasks (tokenization, entity recognition and entity linking) is implemented in a separate module. For tokenization, the spaCy[13] pipeline is used, and the methods used in the rest of the modules are described in the subsections below. The system has a user interface that

[13] https://spacy.io

can be used to input a patient history record and visualize the annotated result and an application programming interface (API), which can be used by other services to consume the annotated results. The annotation results are returned in JSON format.

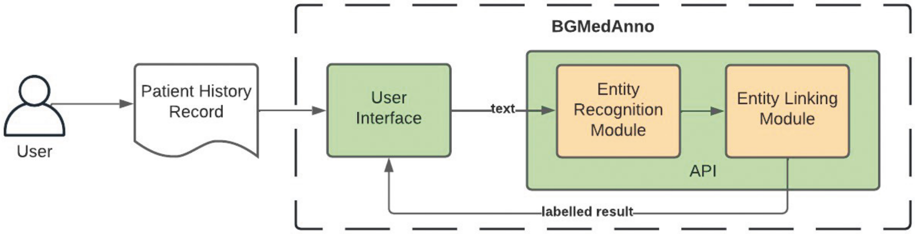


Figure 3: Overall architecture of the tool for automatic annotation BGMedAnno

4.1. Medical terms recognition

For the task of recognizing medical terms, a model is trained using the automatically labeled data in the Patient History Dataset. We use a transfer learning approach based on a model from the BERT family.

BERT (Bi-directional Encoders Representations from Transformers) [15] is a widely used deep learning model based on the transformers architecture. The model is pre-trained on a huge corpus through unsupervised learning and can be further trained on specific tasks like text classification, token classification (named entity recognition), and others. BERT models use context information to generate vector representations of each word.

As a result of the automatic labeling, we have nested entities in 4 different levels. We use MBG-ClinicalBERT by Velichkov et al. [18], a public model based on ClinicalBERT, which was initially trained on 2 million clinical texts in English and further trained on 10,000 medical articles in Bulgarian. We train the MBG-ClinicalBERT-NER model to recognize objects from all four levels at the same time. The standard BERT training architecture for named entity recognition does not support recognizing overlapping entities. For this purpose, we adopt the method of training by using a multi-task learning approach.

Figure 4 shows the architecture of the model for training on multiple tasks. We add four "heads" on top of BERT, i.e., four multi-class classifiers using the standard architecture by Devlin et al. [15] and sharing the same MBG-ClinicalBERT encoder, and we train each classifier on one level of the entities. The different levels of entities are related to each other by rules, for example, "pain in the chest area" has level 1 entities of "pain" and "chest area" and level 3 entity "pain in the chest area". As input for each task, we use all sentences from the training set which contain entities from the task level.

The algorithm for training uses the steps outlined by Liu et al. [19], combines the data for all tasks, and trains the shared encoder on all of them.

4.2. Medical terms linking

Using the terms recognized by the named entity recognition process, the next step is to link them to the corresponding entity in the knowledge base (if it exists). The method for entity linking uses the Medical Term Knowledge Base introduced in Section 3.2.3 . The method uses searching for the closest entity using cosine similarity between the vector representations of the entities in the knowledge base and the input text. The vector representations are generated using MBG-ClinicalBERT as an encoder. For each entity mentioned, the entities with the highest cosine similarity from the knowledge base are identified. If the highest cosine similarity is higher than the threshold (0.8), then the corresponding concept is linked to the term mentioned. If there are no entities with scores higher than the threshold, the entity is linked with the NIL entity.

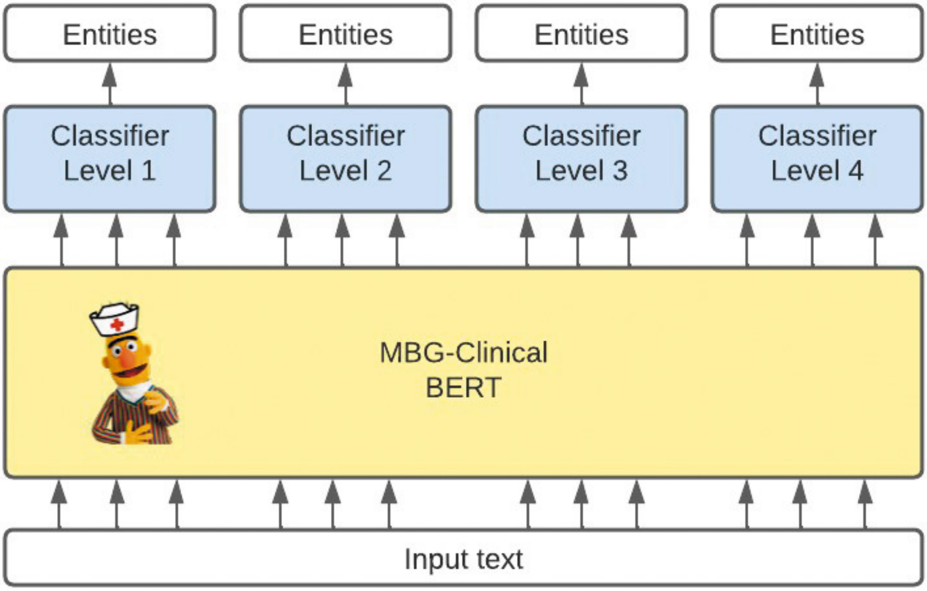


Figure 4: Architecture of the entity recognition model using multi-task learning

Since each word can be represented with one or more tokens using the BERT tokenizer, we use a pooling strategy to combine the vector representations of the tokens making up the word. Devlin et al. [15] compare different pooling strategies like mean, sum, or using the [CLS] token. It is possible to combine different layers from the BERT model to create the token representation – using the second to last layer, combining the last few layers, or concatenating the last few layers. In our paper, we use the sum of the last four layers, so that we use additional feature information from multiple layers but without increasing the vector dimensions which would increase the execution time and memory requirements. For each term in the knowledge base, a vector representation is generated using MBG-ClinicalBERT encoder, summing the last 4 layers of each token, and averaging the token representations to obtain the term representation. The vector representations are stored so that they can be easily used for cosine similarity search during the entity linking process.

5. Experiments for evaluation of the used methods

We conducted experiments to evaluate the performance of the implemented entity recognition and entity linking methods.

5.1. Evaluation of medical terms recognition

For the evaluation of medical terms recognition methods, a small test dataset is manually annotated with nested entities of the different categories. The dataset consists of 100 records. The test dataset is initially labeled using the rules and dictionaries approach and then manually reviewed and fixed. The automatic entity recognition using rules and dictionaries is used as a baseline. We compare the baseline and the trained MBG-ClinicalBERT-NER model using the macro-F1 approach. The evaluation metric F1 uses the exact match of named entities and is calculated as the harmonic mean of precision and recall. Macro-F1 is used as the dataset is highly imbalanced and macro-F1 will average the F1 score for each class. The library *seqeval*[14] is used for evaluation.

The results from the experiment for evaluation of the medical terms recognition experiments for each level are shown in Figure 5. The annotation using rules and dictionaries is doing better on the test set than MBG-ClinicalBERT-NER on all levels except level 2. On average the rules and dictionaries approach is showing a slightly better F1 score – 75%, while MBG-ClinicaBERT-NER has a 73% F1 score. This could be explained by the fact that the test set contains only records in which the automatic annotation has found entities.

Further experiments with different examples, in which the rules and dictionaries approach finds nothing, show that MBG-ClinicalBERT-NER successfully recognizes some entities. For example, in the sentence "Д. панкреатикус е недилатиран." ("ductus pancreaticus is not dilated"), the MBG-ClinicalBERT-

[14] https://huggingface.co/metrics/seqeval

NER model correctly recognizes "ductus pancreaticus" as an organ. A hybrid combination of the two approaches using rules and MBG-ClinicalBERT-NER could recognize more entities and can compensate for the limitations of the dictionaries.

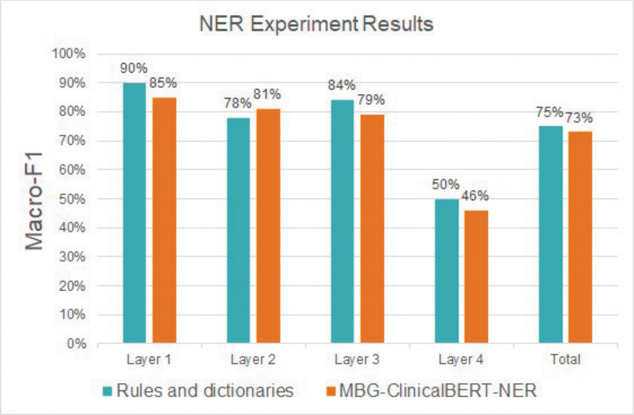


Figure 5: Results of evaluation of the NER model MBG-Clinical-BERT-NER

5.2. Evaluation of medical terms linking

For the evaluation of medical terms linking methods, we use the Outpatient Records Dataset from Section 3.2.4. It was manually annotated with terms, categories, and WikiData concepts. We compare a direct string search in the knowledge base as a baseline, with the developed approach using cosine similarity between vector representations. As evaluation metrics, we use token-based accuracy and F1 score. A token is correctly linked if the model predicts one of the concepts which is true for the token. We calculate precision, recall, and F1 using the library *sklearn.metrics*[15] for the multi-class classification task.

The results from experiments for evaluation of the entity linking method show that the MBG-ClinicalBERT model has an accuracy of 57% and F1-score 61%, while the baseline has an accuracy of 46% and F1-score 45%. We perform several combined experiments with different approaches for entity recognition and linking. The success of the entity linking is limited by the number of correctly recognized terms. The experiments were performed using the Outpatient Records dataset. The results are shown in Table 3. The Outpatient Records dataset is quite different from the dataset used to train MBG-ClinicalBERT-NER, but the model shows a consistent F1-score of 73%, while the rules and dictionaries approach, which was built using similar data to the test, shows 87% F1-score. The entities which were not recognized limit the success of the entity linking and thus the

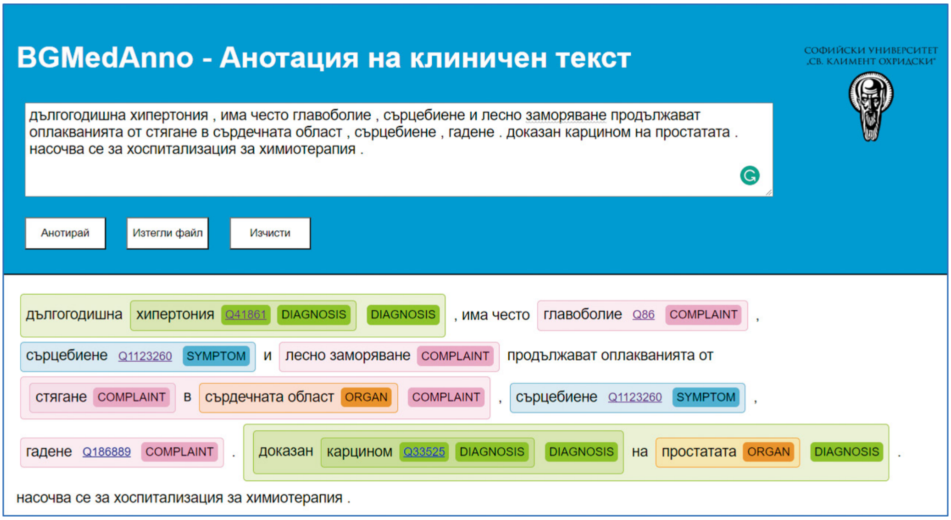
[15] https://scikit-learn.org/stable/index.html

linking results are lower when using MBG-ClinicalBERT-NER – F1-score of 53% compared to 61% when using the rules and dictionaries approach.

Table 3
Combined evaluation results for entity recognition and linking.

Entity Recognition Model	Entity Linking Model	Recognition (F1)	Linking (F1)	Total
Rules and dictionaries	Direct string search	87%	45%	66%
Rules and dictionaries	Linking using MBG-ClinicalBERT	87%	61%	74%
MBG-ClinicalBERT-NER	Direct string search	73%	37%	55%
MBG-ClinicalBERT-NER	Linking using MBG-ClinicalBERT	73%	53%	63%

6. BGMedAnno user interface



BGMedAnno - Анотация на клиничен текст

дългогодишна хипертония , има често главоболие , сърцебиене и лесно заморяване продължават оплакванията от стягане в сърдечната област , сърцебиене , гадене . доказан карцином на простатата . насочва се за хоспитализация за химиотерапия .

дългогодишна хипертония Q41861 DIAGNOSIS DIAGNOSIS , има често главоболие Q86 COMPLAINT ,
сърцебиене Q1123260 SYMPTOM и лесно заморяване COMPLAINT продължават оплакванията от
стягане COMPLAINT в сърдечната област ORGAN COMPLAINT , сърцебиене Q1123260 SYMPTOM ,
гадене Q186889 COMPLAINT . доказан карцином Q33525 DIAGNOSIS DIAGNOSIS на простатата ORGAN DIAGNOSIS .
насочва се за хоспитализация за химиотерапия .

Figure 6: Example of annotated clinical text in Bulgarian with medical terms in the BGMedAnno tool. The annotated text translates into English: *"Long standing hypertension, frequent headaches, palpitations and easy fatigue. continuing complaints of tightening in the chest area, palpitations, nausea, proven prostate carcinoma. redirected for hospitalization for chemotherapy."* The tool recognized 2 diagnoses (highlighted in green), 2 anatomical organs (in orange), 2 symptoms (in cyan), and 4 complaints (in pink). For 6 out of 10 medical concepts, their corresponding WikiData entities are identified, and links are provided

The automatic annotation tool provides an API for clinical text annotation as well as a web user interface, which allows entering a text and visualizing the extracted entities, their categories, and links to the corresponding WikiData concept pages. The API returns the annotated results in JSON format and supports nested entities of different categories. The user interface sends the entered text to the API and visualizes the results in the browser. The user can also download the JSON results as a file for further processing. An example of the user interface with annotated clinical text in Bulgarian is shown in Figure 6.

7. Conclusion

In this paper, we presented a system for automatic annotation of clinical text, named BGMedAnno. The system detects medical terms from the categories of symptoms, complaints, diagnoses, anatomical organs and systems, risk factors, family history, as well as negation of symptoms, complaints, and diagnoses. The system connects the detected entities with the relevant concepts from WikiData, thus linking them to the concepts in UMLS, MeSH, and ICD-10. BGMedAnno detects nested objects and visualizes them.

The paper explores and compares different approaches to discovering and linking medical terms. Methods based on rules and dictionaries as well as based on a trained BERT language model have been developed for the task of finding nested objects. The rule-based method shows an F1 score of 75%, while the trained BERT-based model achieves an F1 score of 73%. Although the BERT model performs slightly worse on the test set, observations show that it finds objects in sentences in which the rule method finds nothing. For the object linking task, the developed method based on the BERT language model shows a 61% F1 result, significantly outperforming direct string comparison, which achieves an F1 result of 45%.

As a future development, additional data can be collected and annotated to improve the results of the recognition model. It is possible to study hybrid models combining rule-based models and deep neural networks, as well as hierarchical models for learning multiple tasks so that the results of lower-level tasks can be used as input for subsequent levels. For the linking task, opportunities can be explored to train a linking model based on automatically annotated data, similar to the recognition approach.

8. Acknowledgments

This research is funded by the Bulgarian Ministry of Education and Science, grant DO1-200/2018 “Electronic health care in Bulgaria” (e-Zdrave) and is also partially supported by Project UNITe BG05M2OP001-1.001-0004 funded by the OP “Science and Education for Smart Growth” and co-funded by the EU through the ESI Funds.

9. References

[1] M. Neves, J. Ševa, An extensive review of tools for manual annotation of documents, Briefings in bioinformatics 22 (2021) 146–163.

[2] J. Li, A. Sun, J. Han, C. Li, A survey on deep learning for named entity recognition, IEEE Transactions on Knowledge and Data Engineering 34 (2020) 50–70.

[3] G. K. Savova, J. J. Masanz, P. V. Ogren, J. Zheng, S. Sohn, K. C. Kipper-Schuler, C. G. Chute, Mayo clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications, Journal of the American Medical Informatics Association 17 (2010) 507–513.

[4] L. Chen, Y. Gu, X. Ji, C. Lou, Z. Sun, H. Li, Y. Gao, Y. Huang, Clinical trial cohort selection based on multi-level rule-based natural language processing system, Journal of the American Medical Informatics Association 26 (2019) 1218–1226.

[5] A. Stubbs, M. Filannino, E. Soysal, S. Henry, Ö. Uzuner, Cohort selection for clinical trials: n2c2 2018 shared task track 1, Journal of the American Medical Informatics Association 26 (2019) 1163–1171.

[6] G. Todorova, Information extraction from medical texts in Bulgarian, Master's thesis, Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski", Sofia, Bulgaria, 2020.

[7] D. Tcharaktchiev, Z. Angelov, S. Boytcheva, G. Angelova, Automatic generation of a national diabetes register from outpatient records, Math. Modeling 2 (2018) 163–166.

[8] S. Zhang, N. Elhadad, Unsupervised biomedical named entity recognition: Experiments with clinical and biological texts, Journal of biomedical informatics 46 (2013) 1088–1098.

[9] P. McNamee, J. Mayfield, Entity extraction without language-specific resources, in: COLING-02: The 6th Conference on Natural Language Learning 2002 (CoNLL-2002), 2002.

[10] A. McCallum, W. Li, Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons (2003).

[11] B. Settles, Biomedical named entity recognition using conditional random fields and rich feature sets, in: Proceedings of the international joint workshop on natural language processing in biomedicine and its applications (NLPBA/BioNLP), 2004, pp. 107–110.

[12] A. Magge, M. Scotch, G. Gonzalez-Hernandez, Clinical ner and relation extraction using bi-char-lstms and random forest classifiers, in: International workshop on medication and adverse drug event detection, PMLR, 2018, pp. 25–30.

[13] Y. Wu, M. Jiang, J. Xu, D. Zhi, H. Xu, Clinical named entity recognition using deep learning models, in: AMIA Annual Symposium Proceedings, volume 2017, American Medical Informatics Association, 2017, p. 1812.

[14] R. Vunikili, H. Supriya, V. G. Marica, O. Farri, Clinical ner using spanish bert embeddings., in: IberLEF@ SEPLN, 2020, pp. 505–511.

[15] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[16] J. Canete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, J. Pérez, Spanish pre-trained bert model and evaluation data, Pml4dc at iclr 2020 (2020) 2020.

[17] S. Boytcheva, B. Velichkov, G. Velchev, I. Koychev, Automatic generation of annotated corpora of diagnoses with icd-10 codes based on open data and linked open data, in: FedCSIS 2020, IEEE, 2020, pp. 163–167.

[18] B. Velichkov, S. Vassileva, S. Gerginov, B. Kraychev, I. Ivanov, P. Ivanov, I. Koychev, S. Boytcheva, Comparative analysis of fine-tuned deep learning language models for ICD-10 classification task for Bulgarian language, in: RANLP 2021, INCOMA Ltd., Held Online, 2021, pp. 1448–1454. URL: https://aclanthology.org/2021.ranlp-1.162.

[19] X. Liu, P. He, W. Chen, J. Gao, Multi-task deep neural networks for natural language understanding, CoRR abs/1901.11504 (2019). arXiv:1901.11504.