

Multilingual Information Extraction in Clinical Texts Using Deep Learning Approaches

Elena Zotova^{1,2}

¹SNLT group at Vicomtech Foundation, Basque Research and Technology Alliance (BRTA), Mikeletegi Pasealekua 57, Donostia/San-Sebastián, 20009, Spain

²Department of Languages and Computer Systems. University of the Basque Country (UPV-EHU) Paseo Manuel de Lardizabal, 1, Donostia/San-Sebastián, 20018, Spain

Abstract

This article briefly describes doctoral thesis research in biomedical natural language processing. The main goal of the research is to contribute to the task of automatic information extraction from unstructured clinical narratives, which includes entity (term, concept) detection and classification, entity linking and clinical coding. Amongst the contributions already made are: a tool for clinical codes mapping and interoperability ClinIDMap, participation in shared tasks dedicated to the clinical named entity recognition, linking and clinical coding where the results of the experiments were published. In addition, the data augmentation method with codes mapping is described.

Keywords

Biomedical NLP, Named Entity Recognition, Entity Linking, Entity Normalisation, Clinical Coding

1. Motivation

My thesis is dedicated to multilingual information extraction in clinical narratives with deep learning approaches. This research aims to contribute to natural language processing (NLP) methods in the clinical domain in Spanish and other languages. The significant growth in Electronic Health Records (EHR) over the last decade has resulted in a rich availability of clinical texts stored in an unstructured format. Much valuable information embedded in the free text can help physicians and medical experts provide better treatment, management and research. Hence, there is a need to explore robust automated techniques to extract such information. In this research, I focus on the automatic clinical coding task that involves detecting a medical concept which occurs in the free text, classifying it, and linking it to a clinical knowledge base (KB); also, the problem of annotated data augmentation is involved.

Various factors are considered important for the growing demand for biomedical NLP, especially for Spanish and other non-English languages. The main of them are the following.

Clinical Data Digitization. The Spanish National Health System has almost completed the digitisation of documents. More than 200,000,000 electronic clinical documents in the National Health System are now in digital format and this number will grow. According to reports,


Doctoral Symposium on Natural Language Processing from the Proyecto ILENIA, 28 September 2023, Jaén, Spain

✉ ezotova@vicomtech.org (E. Zotova)

🆔 0000-0002-8350-1331 (E. Zotova)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

91% of the population of the country has reference to some form of EHR by the year 2021¹. About 80% of a patient's relevant clinical information is written in natural language, such as free text fields of the EHRs, discharge summaries, progress notes, physician's clinical notes, laboratory reports etc. More than 170,000 articles and 28,000 links to the full texts are published in a collection of Spanish scientific journals in the health sciences SciELO España². One of the main tasks derived from the broad use of digital texts is the need for automatic and fast extraction of structured information from unstructured texts.

Difficulty of clinical coding. Clinical coding is a task applied in healthcare administration and consists of assigning medical reports to one or more representative codes. A clinical coder (doctor) analyses a medical report and assigns relevant codes for diagnoses and procedures at the document level. It is a highly costly procedure in terms of time and resources. In addition, special training is needed, and it is obviously manual work. The annual funding of each hospital depends on the ICD-10³ [1] codes reported in all clinical reports. An automatic clinical coding system can support coders to do the job faster and more agile, providing a more specific level of support and relevant detailed information. Using other knowledge bases such as SNOMED-CT [2], UMLS [3] provides extra valuable information for coding medical reports.

Lack of annotated data in Spanish. The data in biomedical NLP is especially difficult to obtain for two reasons. First, clinical reports usually have privacy issues. Even after de-identification, obtaining free access to this kind of medical data is complicated. Therefore, there are very few clinical corpora freely available for research. Second, manual annotation requires high-level expertise, making using crowd-sourcing platforms almost impossible and more expensive than general-purpose NLP-corpora. The problem is even more difficult in multilingual settings since few resources are available for languages other than English. Most existing systems are for English, and few are in production in a real environment.

2. Background

2.1. Clinical Concepts Detection

Named Entity Recognition (NER) [4] in the clinical domain is crucial to extracting concepts (in general-purpose NLP known as named entities) from clinical narratives, such as specific locations, treatment plans, medicines/drugs, diagnoses, etc. Clinical NER is more challenging than general-purpose NER because of three factors. First, entities in clinical texts are nested and ambiguous. Physicians often use abbreviations, acronyms, and synonyms, making standardising difficult. Second, clinical terms can have different meanings, which vary depending on the context. Although this problem mostly applies to non-clinical notes, for clinical NER, this becomes more challenging as the model should understand the complete clinical context along with the entity. A common issue is negative medical findings, where text is written to report findings in a negative context; however, the NER considers that a positive. Third, text spans in clinical texts are long and discontinued.

¹<https://www.sanidad.gob.es/areas/saludDigital/historiaClinicaSNS/mapa/situacionActualHCDSNS.htm>

²<https://scielo.isciii.es/scielo.php>

³<https://icd.who.int/browse10/2019/en>

Clinical NER has the same evolution track as the general-purpose NLP tasks from lexical and rule-based methods to deep neural networks and pre-trained large language models [5]. There is a large variety of BERT-based [6] language models [7] pre-trained on large amounts of texts of the biomedical and clinical domain. Recently, such models for Spanish were created and showed good performance [8]. The primary method for entity recognition is sequence labelling—a task which assigns a class or label to each span of text in a given input sequence.

2.2. Entity Linking

Entity Linking (EL), or entity normalisation, is the key technology enabling semantic applications and informatics pipelines in the biomedical domain. EL is the task of establishing a link between a concept detected in the unstructured text to an entry in a structured knowledge graph/database. A popular and fast method is exact string matching against a database of synonyms; its advantage is high precision while it suffers from low recall [9]. A rule-based algorithm [10] which applies various transformations such as stemming, suffix replacement, and acronym expansion, also achieved high precision but struggle with tasks requiring softer reasoning.

Recently, the methods are based on Semantic Text Similarity (STS) pipelines [11]. Semantic EL system consists of 1) entity and context encoding, 2) candidate generation, and 3) candidate ranking. The entity encoders have shifted to self-attention architectures and started using deep pre-trained models like BERT [12]. Most studies rely on external knowledge for the candidate generation step. There is a surge of models that tackle the domain adaptation problem in a zero-shot fashion. The learning type for the disambiguation can be supervised, unsupervised, weakly supervised, or zero-shot. Some models are trained on cross-lingual data [13]. The main challenge of EL in the clinical domain is the large variety of synonyms in clinical ontologies, their hierarchical structure and the rich context where they occur.

2.3. Existing Tools

Most clinical concept detection and linking tools are for English and link mentions to UMLS concepts. Some are also private and developed by big companies like Amazon or IBM. cTAKES[3] uses a dictionary look-up and each mention is mapped to a UMLS concept. MetaMap⁴ is a tool that identifies UMLS concepts in the text of clinical. It is based on a lexical lookup of input words. CLAMP [14] takes two approaches: a machine learning using Conditional Random Field and a dictionary-based approach, which maps mentions to standardised ontologies. Spark NLP⁵ and Amazon Comprehend Medical⁶ map clinical findings to ICD-10-CM, SNOMED CT, RxNorm and other codes. Kodifica⁷ for Spanish is rule-based and dictionary-based.

⁴https://www.nlm.nih.gov/research/umls/implementation_resources/metamap.html

⁵https://demo.johnsnowlabs.com/healthcare/ER_ICD10_CM/

⁶<https://aws.amazon.com/es/comprehend/medical/>

⁷<https://ibermatica.com/wp-content/uploads/sites/4/2017/11/Kodifica-T-Codif.-sem%C3%A1ntico-de-diagn%C3%B3stico.pdf>

3. Hypothesis and Research Proposal

This thesis hypothesises that it is possible to automatically extract structured information with high accuracy from unstructured texts written in natural language. To prove the hypothesis, the following tasks and experiments are proposed.

- Develop and evaluate automatic systems based on state-of-the-art algorithms for extracting structured information from medical records, including clinical entity detection and linking them to clinical KBs for non-English languages.
- Experiment with the NER addressing the most challenging points, such as concept disambiguation, long-span detection, and multiclass classification.
- Experiment with EL techniques, using semantic similarity and text generation techniques.
- Create a tool prototype for semantic interoperability of clinical concepts among KBs and lexical resources.
- Experiment with automatic and semiautomatic methods to create new annotated data.

4. Methodology and Experiments

4.1. Clinical codes mapping for data augmentation

We research data augmentation methods in the situation of scarce annotated data. One of the first steps in our research is an application ClinIDMap [15, 16], which is aimed to map clinical codes of taxonomies (UMLS, ICD-10, MeSH, SNOMED-CT etc.) and lexical resources (Wikipedia, WordNet). We created it to make the clinical codes interoperable and use for generating more annotated data. Most clinical concepts are transferable across various knowledge bases and languages. The alignment uses the IDs of the KBs from the official mapping resources developed by SNOMED-CT and UMLS authors.

The alignment allows us to enrich manually annotated corpora with extra clinical codes. If we have a corpus annotated in UMLS codes, we can map each code to ICD-10-CM and ICD-10-PSC codes to automatically derive a new version of the corpus for training a new SL system. And vice versa, a corpus annotated with ICD-10 codes can be used to automatically derive corpora annotated with UMLS codes, semantic types or groups. Our tool also enriches the concepts with multilingual terms and descriptions of its available Wikipedia articles, which allows us to expand brief taxonomy descriptions to detailed information in multiple languages. For instance, a Spanish sentence below annotated with a UMLS code C0011860 can be mapped to SNOMED CT code 44054006, ICD-10-CM code E11.9, semantic group Disorder and the corresponding Wikipedia articles in 51 languages.

La paciente presentaba como antecedentes personales hipertensión y **diabetes tipo 2** (C0011860) controladas mediante tratamiento médico convencional. (*The patient had a history of hypertension and **type 2 diabetes** controlled by conventional medical treatment.*)

We experiment with the following corpora. **CodiEsp 2020** [17] a corpus of clinical cases, manually curated by the CLEF e-Health shared task organisers. It is annotated with Diagnosis

(*diagnóstico*) or a Proceeding (*procedimiento*) and ICD-10 codes. **E3C Corpus** [18] is a multilingual corpus in English, French, Italian, Spanish, and Basque of clinical narratives annotated with semantic groups (e.g., pathologies, drugs, anatomy, etc.) and temporal information and factuality (e.g., events). **CT-EBM-SP** [19] is a collection of texts in Spanish annotated with UMLS codes and semantic groups. **MANTRA** corpus [20] consists of parallel corpora in English, French, German, Spanish, and Dutch manually annotated with the biomedical concepts and UMLS codes. **MedMentions** [21] is a large English dataset of PubMed abstracts annotated with UMLS concepts.

Using the mapping tool, we derive multiple datasets from existing datasets annotated with different coding systems and obtain a new larger corpus. The resulting annotated corpus is prepared to train NER models to classify semantic groups, diagnoses, and procedures. We map ICD-10 categories to UMLS Semantic groups and vice versa. Finally, we compare classification models trained on the gold-standard corpus and corpus annotated with the mapping method. We train several deep learning models and see that the models trained on the corpora annotated with our method, perform quite well comparably with the gold standard (See the results in [15]). The tool's code is publicly available⁸.

4.2. Entity Linking Experiments

For the EL task, our primary approach is STS techniques, where two texts' degree of semantic closeness is measured. Semantic search is based on STS, allowing retrieval of relevant text results beyond mere lexical matching. The main concepts of semantic search are query, collection of documents, and level of relevance between a query and documents. It typically involves embedding all documents (sentences or, in this case, clinical taxonomy descriptions) into a vector space, this process is also known as encoding. The query, represented by the detected entity (as described in Subsection 2.1), is embedded into the same vector space at search time. Nowadays, the most extended method to encode text is to use a pre-trained Transformer model [22] to obtain the corresponding embeddings (multidimensional vectors). These models are used as encoders. Then, we compute the score using a similarity metric. The metric can be cosine distance, dot product and others. The top N documents with the shortest distance or with the higher scores is the set of candidates for linking.

We experiment with different types of encoders, trained on the domain data, for example, SapBERT-XLMR-large model [23]. This model is trained with UMLS. We find injecting UMLS knowledge of multilingual clinical terminology into a pre-trained language model especially helpful for the normalisation task. Next, each corpus entity's closest candidate from the SNOMED CT is retrieved. The code of the most similar taxonomy entry is used as the predicted code for each given entity.

We also experiment with a cross-encoder model [24] training. Cross-encoders handle sentence pair scoring and classification tasks [25]. They have been proven successful in the clinical domain also [26]. In contrast to an unsupervised semantic similarity function, the cross-encoder is trained by encoding a pair of sentences simultaneously and producing a value between 0 and 1 that indicates the similarity or relatedness of the input sentence pair. Cross-encoders are

⁸<https://github.com/Vicomtech/ClinIDMap>

trained using a set of text pairs labelled as similar/related (i.e., positive) or dissimilar/unrelated (negative). This method allows us to fine-tune the model on the annotated domain data. After retrieving of top N candidates with the semantic similarity method, we can rerank the document with the cross-encoder model predicting the score for each query-document pair.

4.3. Participation in Shared Tasks

In order to prove the hypothesis and obtain access to annotated data, we have participated in various shared tasks organised during IberLEF⁹ or CLEF¹⁰ conferences.

MEDDOPROF 2021 [27]. The task was to detect mentions of professions in the clinical domain. In the normalisation task, which is the same as entity linking, the participants had to link the entities with SNOMED CT, assigning the code to each entity. Our results were better than the baseline in the NERC and classification tasks. In normalisation, we did not pass the baseline. We presented a joint model based on BERT, and for the normalisation task, we used semantic similarity to search for the closest candidate in the vector space [28].

LivingNER 2022 [29]. The task had three tracks: detection of entities meaning living beings, link them to a knowledge base, in this case, NCBI¹¹, a large taxonomy of biomedical information, and the third track was to classify the detected entities as a pet, an animal injury, food or a nosocomial infection. We have presented the models based on transformer models. In the entity linking task we experimented with vector space and methods of linking semantically identical texts. Our system gets the best result in task 2 [30].

MedProcNER 2023 [31] The task is designed similarly, where the first subtask is to detect entities, in this case, procedures, and the second subtask is entity linking to SNOMED CT taxonomy. The third task is clinical indexing, where the system should detect all SNOMED CT codes in a given document. The results are to be published in September 2023 at CLEF conference.

4.4. Entity Linking as Text Generation Task

As was described in Subsection 4.2, the primary approach for EL task is a semantic search for the closest candidates and candidates ranking. Recently, the sequence to sequence (seq2seq) approach [32] is trending in the NLP, which consists of taking the text as input and producing new text as output. Our experiment is to train models capable generate taxonomy definitions from the corpus concepts; for example, taken as inputs, corpus words “dolor ótico” (ear pain) can generate the output “otalgia, oído no especificado” (otalgia, ear not specified), which is code H92.09 in ICD-10. We want to experiment with seq2seq models, such as T5 [32], to determine if this new approach can improve previous work on EL.

⁹<http://sepln2023.sepln.org/iberlef/>

¹⁰<http://www.clef-initiative.eu/>

¹¹<https://www.ncbi.nlm.nih.gov/>

5. Research Elements for Discussion

My research is in the middle of the way. There are still many questions to address and problems to resolve. I would like to list some of them, the most challenging from my point of view.

1. **Data augmentation.** As was said above, there is a lack of annotated data for clinical NLP. How do we get more corpora in the situation of highly sensitive data and the high cost of manual annotation? Which methods are applicable? How can we generate more multilingual corpora?
2. **Challenges for clinical entity recognition.** As was said in Subsection 2.1, clinical concepts are much more difficult to detect and classify, because of a large variety of lexical representation, their nested nature and their high impact of context, compared to classical location–name–organisation setting. Is the classical NER approach the best?
3. **Entity Linking task** is far from being resolved in clinical narratives, and the performance of even the state-of-the-art systems highly depends on the quality and nature of the training data. It is even more difficult to annotate codes than entities and concepts. There are a lot of situations of ambiguity when the same lexical representation can be related to various codes depending on the context and other factors. Clinical dictionaries and taxonomies are more specific than other lexical resources such as Wikipedia. Which types of semantic similarity and ranking models perform best? Do we need to encode the concept detected in the text only or use its context, too? Can we use other than semantic similarity techniques, such as text generation to implement entity linking system?

Acknowledgments

This doctoral thesis is supported by Vicomtech Foundation. I am grateful to my thesis supervisors German Rigau (EHU/UPV) and Montse Cuadros (Vicomtech) for their guidance and help during the work done up to now.

References

- [1] World Health Organization (WHO), ICD-10 : international statistical classification of diseases and related health problems : tenth revision, 2nd ed ed., World Health Organization, 2004.
- [2] K. Donnelly, et al., SNOMED-CT: The advanced terminology and coding system for eHealth, *Studies in health technology and informatics* 121 (2006) 279.
- [3] O. Bodenreider, The Unified Medical Language System (UMLS): integrating biomedical terminology, *Nucleic Acids Res.* 32 (2004) 267–270.
- [4] B. Jehangir, S. Radhakrishnan, R. Agarwal, A survey on Named Entity Recognition – Datasets, Tools, and Methodologies, *Natural Language Processing Journal* 3 (2023) 100017.
- [5] P. Bose, S. Srinivasan, W. C. Sleeman, J. Palta, R. Kapoor, P. Ghosh, A Survey on Recent Named Entity Recognition and Relationship Extraction Techniques on Clinical Texts, *Applied Sciences* 11 (2021).

- [6] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019, pp. 4171–4186.
- [7] K. S. Kalyan, A. Rajasekharan, S. Sangeetha, AMMU: A Survey of Transformer-Based Biomedical Pretrained Language Models, *J. of Biomedical Informatics* 126 (2022).
- [8] A. G. Fandiño, J. A. Estapé, M. Pàmies, J. L. Palao, J. S. Ocampo, C. P. Carrino, C. A. Oller, C. R. Penagos, A. G. Agirre, M. Villegas, MarIA: Spanish Language Models, *Procesamiento del Lenguaje Natural* 68 (2022).
- [9] C. Jonquet, N. Shah, C. Youn, M. Musen, C. Callendar, M.-A. Storey, NCBO Annotator: Semantic Annotation of Biomedical Data, *ISWC* (2009).
- [10] J. D’Souza, V. Ng, Sieve-Based Entity Linking for the Biomedical Domain, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Association for Computational Linguistics, Beijing, China, 2015, pp. 297–302.
- [11] S. Kim, N. Fiorini, W. J. Wilbur, Z. Lu, Bridging the Gap: Incorporating a Semantic Similarity Measure for Effectively Mapping PubMed Queries to Documents, *Journal of Biomedical Informatics* 75 (2017) 122–127.
- [12] N. Reimers, I. Gurevych, Making Monolingual Sentence Embeddings Multilingual Using Knowledge Distillation, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2020.
- [13] O. Sevgili, A. Shelmanov, M. Arkhipov, A. Panchenko, C. Biemann, Neural Entity Linking: A Survey of Models Based on Deep Learning, *Semantic Web* 13 (2022) 1–44.
- [14] E. Soysal, J. Wang, M. Jiang, Y. Wu, S. V. S. Pakhomov, H. Liu, H. Xu, CLAMP – a toolkit for efficiently building customized clinical natural language processing pipelines, *Journal of the American Medical Informatics Association : JAMIA* 25 (2017) 331 – 336.
- [15] E. Zotova, M. Cuadros, G. Rigau, ClinIDMap: Towards a Clinical IDs Mapping for Data Interoperability, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 3661–3669.
- [16] E. Zotova, Towards the integration of WordNet into ClinIDMap, *Global WordNet Conference*, 2023.
- [17] A. Miranda-Escalada, A. Gonzalez-Agirre, J. Armengol-Estapé, M. Krallinger, Overview of automatic clinical coding: annotations, guidelines, and solutions for non-english clinical cases at CodiEsp track of CLEF eHealth 2020 (2020).
- [18] B. Magnini, B. Altuna, A. Lavelli, M. Speranza, R. Zanolì, The E3C Project: Collection and Annotation of a Multilingual Corpus of Clinical Cases (2020).
- [19] L. Campillos-Llanos, First Steps towards Building a Medical Lexicon for Spanish with Linguistic and Semantic Information (2019) 152–164.
- [20] J. A. Kors, S. Clematide, S. A. Akhondi, E. M. van Mulligen, D. Rebholz-Schuhmann, A multilingual gold-standard corpus for biomedical concept recognition: the Mantra GSC, *Journal of the American Medical Informatics Association* 22 (2015) 948–956.
- [21] S. Mohan, D. Li, MedMentions: A Large Biomedical Corpus Annotated with UMLS Concepts, *ArXiv abs/1902.09476* (2019).

- [22] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, I. Polosukhin, Attention is All you Need, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, volume 30, Curran Associates, Inc., 2017.
- [23] F. Liu, I. Vulić, A. Korhonen, N. Collier, Learning Domain-Specialised Representations for Cross-Lingual Biomedical Entity Linking, in: *Proceedings of ACL-IJCNLP 2021*, 2021, pp. 565–574.
- [24] N. Reimers, I. Gurevych, Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, in: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Association for Computational Linguistics, Hong Kong, China, 2019, pp. 3982–3992.
- [25] R. Nogueira, K. Cho, Passage Re-ranking with BERT, *ArXiv abs/1901.04085* (2019).
- [26] A. Rahimi, T. Baldwin, K. Verspoor, WikiUMLS: Aligning UMLS to Wikipedia via Cross-lingual Neural Ranking, in: *Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online)*, 2020, pp. 5957–5962.
- [27] S. Lima-López, E. Farré-Maduell, A. Miranda-Escalada, V. Brivá-Iglesias, M. Krallinger, Nlp applied to occupational health: Meddoprof shared task at iberlef 2021 on automatic recognition, classification and normalization of professions and occupations from medical texts, *Procesamiento del Lenguaje Natural 67* (2021) 243–256.
- [28] E. Zotova, A. García-Pablos, M. Cuadros, Vicomtech at MEDDOPROF: automatic information extraction and disambiguation in clinical text, in: *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021) co-located with the Conference of the Spanish Society for Natural Language Processing (SEPLN 2021), XXXVII International Conference of the Spanish Society for Natural Language Processing.*, Málaga, Spain, September, 2021, volume 2943 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 776–787.
- [29] A. Miranda-Escalada, E. Farré-Maduell, S. Lima-López, D. Estrada, L. Gascó, M. Krallinger, Mention detection, normalization & classification of species, pathogens, humans and food in clinical documents: Overview of livingner shared task and resources, *Procesamiento del Lenguaje Natural* (2022).
- [30] E. Zotova, A. García-Pablos, N. Pérez, P. Turón, M. Cuadros, Vicomtech at LivingNER2022, in: *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022) co-located with the Conference of the Spanish Society for Natural Language Processing (SEPLN 2022)*, A Coruña, Spain, September 20, 2022, volume 3202 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022.
- [31] S. Lima-López, E. Farré-Maduell, L. Gascó, A. Nentidis, A. Krithara, G. Katsimpras, G. Paliouras, M. Krallinger, Overview of medprocner task on medical procedure detection and entity linking at bioasq 2023, in: *Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum*, 2023.
- [32] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. J. Liu, Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer, *Journal of Machine Learning Research* 21 (2020) 1–67.