SKET: an Unsupervised Knowledge Extraction Tool to Empower Digital Pathology Applications*

Extended Abstract

Giorgio Maria Di Nunzio[1], Nicola Ferro[1], Fabio Giachelle[1], Ornella Irrera[1], Stefano Marchesin[1,*] and Gianmaria Silvello[1]

[1]*Department of Information Engineering, University of Padua*

Abstract

Large volumes of medical data have been produced for decades. These data include diagnoses, which are often reported as free text, thus encoding medical knowledge that is still largely unexploited. To decode the medical knowledge present within reports, we propose the Semantic Knowledge Extractor Tool (SKET), an unsupervised knowledge extraction system combining a rule-based expert system with pre-trained Machine Learning (ML) models. This work demonstrates the viability of unsupervised Natural Language Processing (NLP) techniques to extract critical information from cancer reports, opening opportunities such as data mining for knowledge extraction purposes, precision medicine applications, structured report creation, and multimodal learning.

Keywords

Knowledge Extraction, Machine Learning, Expert Systems, Digital Pathology

1. Introduction

Hundred of thousands of medical reports have been used to communicate diagnoses, encoding a vast amount of medical knowledge. In this context, free-text reporting is the de facto standard to communicate diagnoses, guiding patients' treatment, and conducting therapies. Processing high volumes of free-text reports to extract the crucial knowledge is usually performed manually. However, since reports vary widely between institutions, contain noise, and lack a standard structure, this becomes an extremely time-consuming process. To overcome this limitation, Natural Language Processing (NLP) methods become essential [2, 3, 4, 5, 6, 7, 8, 9] as they empower the efficient automatic processing of thousands of reports and the extraction of relevant information for several (downstream) tasks, such as clinical note mining [10, 11] and structuring [12], risk prediction [13], clinical decision support [14], and precision medicine retrieval [15].

In the context of digital pathology – a field that involves the analysis of histopathology images known as Whole Slide Images (WSIs) – this work aims at proving the viability of unsupervised

19th IRCDL (The Conference on Information and Research science Connecting to Digital and Library science), February 23–24, 2023, Bari, Italy

*The full paper has been originally published in the Journal of Pathology Informatics [1]

*Corresponding author.

stefano.marchesin@unipd.it (S. Marchesin)

NLP techniques to automatically extract critical information from pathology reports and use it for different applications, such as automatic report annotation and visualization [16], as well as WSI classification [17]. To this end, we present the Semantic Knowledge Extractor Tool (SKET), an unsupervised hybrid knowledge extraction system that combines rule-based techniques with pre-trained Machine Learning (ML) models to extract knowledge from pathology reports. In recent years, NLP has shifted from using rules to ML approaches [18, 9], which have the advantage of learning regularities from data and of generalizing to previously unseen patterns. Moreover, the advent of efficient Neural Language Models (NLMs) [19, 20, 21, 22] paved the way for the pre-training era, where large NLMs trained in a self-supervised fashion on huge datasets are used to develop NLP models for a number of downstream tasks. Nevertheless, similarly to [10], we argue that rule-based techniques capture critical information that should be used together with – and not substituted by – ML to improve performance.

We evaluate SKET effectiveness on entity linking and text classification, considering three use-cases: Colon, Cervix, and Lung cancer. We resort on diagnostic reports coming from two medical centers based in Italy and The Netherlands. Then, we compare SKET with unsupervised ML approaches to understand the impact that combining rule-based techniques and pre-trained ML models have on the extraction of knowledge from diagnostic reports. The results highlight the effectiveness of ML methods for information extraction in the pathology domain but, at the same time, they also stress the role of expert knowledge in reaching the high levels of accuracy required to semi-automate the clinical practice. As further proof, SKET has been already used as core system in automatic report annotation and visualization [16], as well as weak supervision for WSI classification [17]. SKET source code is publicly available at https://github.com/ExaNLP/sket.

The rest of this paper is organized as follows: Section 2 presents SKET. Section 3.2 describes the experimental evaluation. Finally, Section 4 concludes the paper.

2. The Semantic Knowledge Extractor Tool

SKET combines pre-trained Named Entity Recognition (NER) models with unsupervised Entity Linking (EL) methods to extract relevant entities from diagnostic reports and link them to concepts stored in a reference ontology[1]. By relying on pre-trained NER models and unsupervised EL methods, SKET can serve as automated annotator in weak supervision tasks. For instance, the concepts extracted by SKET can be used as weak labels when training ML models for image classification [23, 24] and relation extraction [25], or as nodes to build knowledge graphs that can be used for retrieval tasks [26].

SKET consists of four main components: (1) Named Entity Recognition, (2) Entity Linking, (3) Data Labeling, and (4) Graph Creation. Components (1) and (2) are sequential, whereas (3) and (4) can be applied in parallel. We briefly describe each component below.

[1] https://w3id.org/examode/ontology/

2.1. Named Entity Recognition

NER can be defined as the task of identifying and categorizing relevant information within text. A named entity can be any word or phrase – i.e., a mention – that consistently refers to a concept or object of the world. Once identified, mentions are classified into predefined categories, such as disease, gene/protein, symptom, etc.

To perform NER, SKET combines pre-trained neural models with rule-based techniques. As neural component, SKET exploits ScispaCy models [27], which provide full NER pipelines for biomedical data, consisting of large medical vocabularies, as well as Word2Vec [19] word vectors trained on the PubMed Central Open Access Subset [28]. Regarding the integration of expert rules, SKET extends the ScispaCy pipeline with two more components: Entity Fusion and Negation Detection. For **Entity Fusion**, SKET exploits expert rules to identify and merge specific mentions that would otherwise be regarded as separate by ScispaCy. For example, "high-grade" and "dysplasia" are considered as separate mentions, whereas we are interested in "high-grade dysplasia" as a unique mention. Hence, we developed regular expressions capable of identifying trigger terms that are indicative of a set of mentions that should potentially be combined into one. These expert rules have been developed on a holdout dataset, which is available in the SKET GitHub repository[2]. The dataset consists of 50 diagnostic reports for each use-case and medical center, for a total of 250 diagnostic reports. For **Negation Detection**, SKET relies on NegEx [29], a negation detection algorithm that evaluates whether extracted entities are negated within text. NegEx uses regular expressions to identify the scope of trigger terms that are indicative of negation. Then, the entities extracted within the scope of a trigger term are marked as negated and removed.

2.2. Entity Linking

EL can be defined as the task of assigning unique meanings to entities mentioned within text. In a nutshell, EL aims to determine whether a target named entity refers to a specific concept or object stored within a reference ontology.

To perform EL, SKET adopts ad-hoc and similarity-based matching. Given an extracted entity, SKET performs a two-stage matching approach. First, the system tries to link the entity using ad-hoc matching. Then, if ad-hoc matching fails, it employs the similarity-based matching. For **Ad-Hoc Matching**, SKET employs regular expressions to find trigger terms indicative of a specific concept in the ontology. Once a trigger is found, the system matches the entity containing the trigger term with the closest ontology concept. In this case, if an extracted entity contains the (trigger) term "carcinoma", then SKET links the entity to the "colon adenocarcinoma" concept. Ad-hoc matching rules have also been developed on the holdout dataset and are available on GitHub. Regarding **Similarity Matching**, SKET combines string and semantic matching techniques. For string matching, SKET adopts the Gestalt Pattern Matching (GPM) algorithm [30]. For semantic matching, SKET exploits the word vectors provided by ScispaCy models [27]. Specifically, it computes the cosine distance between the vector representations of extracted entities and ontology concepts.

[2] https://github.com/ExaNLP/sket/tree/main/sket/nerd/rules/

2.3. Data Labeling

Given the set of concepts extracted from each diagnostic report, SKET maps a clinically relevant subset of such concepts to a set of annotation classes defined by pathologists.

2.4. Graph Creation

SKET builds report-level knowledge graphs using the extracted concepts as nodes and the semantic relations of the reference ontology as edges. The use of ontology concepts and relations to describe diagnostic reports increases the semantic understanding of the underlying data [31]. Once created, report-level knowledge graphs are encoded in a machine-readable format through RDF.

3. Experimental Evaluation

3.1. Setup

Tasks: We evaluate SKET on Entity Linking (Task 1) and Text Classification (Task 2). Both tasks are addressed as multi-label classification problems. Note that the number of possible labels for entity linking is much higher than for text classification, making the task an extreme multi-label classification problem [32, 33].
Datasets: For Task 1, we use 1,250 annotated reports coming from both medical centers and related to all the three use-cases. For Task 2, we resort on 9,798 annotated reports, divided among medical centers and use-cases. We refer the reader to the original publication [1] for a comprehensive description of the available data.
Baselines: For both tasks, we compare SKET with two unsupervised approaches based on Bio FastText [20, 34] and BioClinical BERT [22, 35]. For a fair comparison, both approaches adopt the same NER ScispaCy pipeline used by SKET, but without the extensions introduced with it. Then, they perform EL by computing the cosine distance between the vector representations of the extracted entities and the ontology concepts. Both baselines are straightforward approaches to perform entity linking and text classification without annotated data.

3.2. Results

Table 1 reports the results obtained by SKET and the considered baselines on Entity Linking (left) and Text Classification (right).

For entity linking (Task 1), we observe that SKET achieves high performance for both micro- and weighted-average F1 in each considered use-case. Regarding accuracy, its performance varies depending on the use-case – with the lowest score obtained in colon cancer with a value of 0.6280. As for the comparison of SKET with the considered baselines, we see that it outperforms them in each use-case for all measures. This result shows the effectiveness of combining ad-hoc, expert rules with ML models – making SKET both precise and sensitive. Specifically, ad hoc matching makes SKET precise, while semantic matching makes it sensitive. To support this intuition, we observe that unsupervised baselines – which only rely on ML models and semantic matching – have low accuracy values. Since we tackle the entity linking task as a multi-label

Table 1
Entity linking (left) and text classification (right) results on colon, cervix, and lung cancer pathology reports. **Bold** values represent the highest scores achieved for each measure.

Entity Linking			
Colon			
Model	Accuracy	Micro F1	Weighted F1
SKET	**0.6280**	**0.8861**	**0.8694**
FastText	0.0660	0.5000	0.6146
BERT	0.1840	0.3905	0.4527
Cervix			
Model	Accuracy	Micro F1	Weighted F1
SKET	**0.7020**	**0.8322**	**0.8368**
FastText	0.0900	0.2802	0.3439
BERT	0.0720	0.2715	0.2940
Lung			
Model	Accuracy	Micro F1	Weighted F1
SKET	**0.8624**	**0.9375**	**0.9262**
FastText	0.2510	0.5610	0.6506
BERT	0.3806	0.6804	0.8395

Text Classification			
Colon			
Model	Accuracy	Micro F1	Weighted F1
SKET	**0.7525**	**0.8386**	**0.8373**
FastText	0.4146	0.5298	0.5514
BERT	0.5167	0.5697	0.6587
Cervix			
Model	Accuracy	Micro F1	Weighted F1
SKET	**0.5281**	**0.7791**	**0.7611**
FastText	0.2533	0.4882	0.4445
BERT	0.3066	0.3962	0.4867
Lung			
Model	Accuracy	Micro F1	Weighted F1
SKET	0.8137	0.8387	0.8262
FastText	0.5221	0.7296	0.6853
BERT	**0.8523**	**0.8630**	**0.8526**

classification problem, we resort on subset accuracy, where the set of concepts predicted for a report must exactly match the corresponding set of ground-truth concepts. Therefore, accuracy values are prone to rapidly decrease and less precise models are naturally affected by this.

For text classification (Task 2), we see that SKET performs well on colon and lung cancer use-cases, whereas it shows lower accuracy values on cervix cancer. This result suggests that the cervix use-case is harder than the others, as subset accuracy drops fast when a model fails to predict all labels correctly. The higher values for micro- and weighted-average F1 – which do not perform exact match between predicted and ground-truth labels – further support this intuition. Compared to baselines, SKET outperforms them in colon and cervix use-cases. On the other hand, the BERT-based approach proves more effective in lung cancer. Despite this, the robustness of SKET across different use-cases makes it a viable solution in real scenarios, where annotated data are hard and expensive to get.

4. Conclusion

In this work, we presented SKET, an unsupervised hybrid knowledge extraction system that combines rule-based techniques with pre-trained ML models to extract relevant concepts from diagnostic reports. The experimental evaluation demonstrated the effectiveness of SKET, making it a viable solution to reduce pathologists' workload. Besides, the experimental results highlighted the importance of expert knowledge in developing unsupervised systems for specialized medicine. As a result, the extracted concepts can serve different digital pathology

applications, such as automatic report annotation, visualization, and retrieval, as well as image classification.

Acknowledgments

The work was supported by the ExaMode project, as part of the EU H2020 program under Grant Agreement no. 825292.

References

[1] S. Marchesin, F. Giachelle, N. Marini, M. Atzori, S. Boytcheva, G. Buttafuoco, F. Ciompi, G. M. Di Nunzio, F. Fraggetta, O. Irrera, H. Müller, T. Primov, S. Vatrano, G. Silvello, Empowering digital pathology applications through explainable knowledge extraction tools, Journal of Pathology Informatics 13 (2022) 100139. doi:https://doi.org/10.1016/j.jpi.2022.100139.

[2] T. Davenport, R. Kalakota, The Potential for Artificial Intelligence in Healthcare, Future Healthc J. 6 (2019) 94–98. URL: https://doi.org/10.7861/futurehosp.6-2-94. doi:10.7861/futurehosp.6-2-94.

[3] J. M. Buckley, S. B. Coopey, J. Sharko, F. Polubriaginof, B. Drohan, A. K. Belli, E. M. Kim, J. E. Garber, B. L. Smith, M. A. Gadd, M. C. Specht, C. A. Roche, T. M. Gudewicz, K. S. Hughes, The Feasibility of Using Natural Language Processing to Extract Clinical Information from Breast Pahology Reports, J. Pathol Inform 3 (2012) 23. URL: https://doi.org/10.4103/2153-3539.97788. doi:10.4103/2153-3539.97788.

[4] S. Hassanpour, C. P. Langlotz, Information Extraction from Multi-Institutional Radiology Reports, Artif. Intell. Medicine 66 (2016) 29–39. URL: https://doi.org/10.1016/j.artmed.2015.09.007. doi:10.1016/j.artmed.2015.09.007.

[5] G. Burger, A. Abu-Hanna, N. de Keizer, R. Cornet, Natural Language Processing in Pathology: a Scoping Review, Journal of Clinical Pathology 69 (2016) 949–955. URL: https://doi.org/10.1136/jclinpath-2016-203872. doi:10.1136/jclinpath-2016-203872.

[6] M. Topaz, L. Murga, K. M. Gaddis, M. V. McDonald, O. Bar-Bachar, Y. Goldberg, K. H. Bowles, Mining Fall-Related Information in Clinical Notes: Comparison of Rule-Based and Novel Word Embedding-Based Machine Learning Approaches, J. Biomed. Informatics 90 (2019). URL: https://doi.org/10.1016/j.jbi.2019.103103. doi:10.1016/j.jbi.2019.103103.

[7] T. Oliwa, S. B. Maron, L. M. Chase, S. Lomnicki, D. V. T. Catenacci, B. Furner, S. L. Volchenboum, Obtaining Knowledge in Pathology Reports Through a Natural Language Processing Approach With Classification, Named-Entity Recognition, and Relation-Extraction Heuristics, JCO Clinical Cancer Informatics 1 (2019) 1–8. URL: https://doi.org/10.1200/CCI.19.00008. doi:10.1200/CCI.19.00008.

[8] K. Kreimeyer, M. Foster, A. Pandey, N. Arya, G. Halford, S. F. Jones, R. Forshee, M. Walderhaug, T. Botsis, Natural Language Processing Systems for Capturing and Standardizing Unstructured Clinical Information: A Systematic Review, J. Biomed. Informatics 73 (2017) 14–29. URL: https://doi.org/10.1016/j.jbi.2017.07.012. doi:10.1016/j.jbi.2017.07.012.

[9] Y. Wang, L. Wang, M. Rastegar-Mojarad, S. Moon, F. Shen, N. Afzal, S. Liu, Y. Zeng,

S. Mehrabi, S. Sohn, H. Liu, Clinical Information Extraction Applications: A Literature Review, J. Biomed. Informatics 77 (2018) 34–49. URL: https://doi.org/10.1016/j.jbi.2017.11.011. doi:10.1016/j.jbi.2017.11.011.

[10] E. Santus, T. Schuster, A. M. Tahmasebi, C. Li, A. Yala, C. R. Lanahan, P. Prinsen, S. F. Thompson, S. Coons, L. Mynderse, R. Barzilay, K. Hughes, Exploiting Rules to Enhance Machine Learning in Extracting Information From Multi-Institutional Prostate Pathology Reports, JCO Clinical Cancer Informatics (2020) 865–874. URL: https://doi.org/10.1200/CCI.20.00028. doi:10.1200/CCI.20.00028.

[11] Y. Kim, J. H. Lee, S. Choi, J. M. Lee, J. H. Kim, J. Seok, H. J. Joo, Validation of Deep Learning Natural Language Processing Algorithm for Keyword Extraction from Pathology Reports in Electronic Health Records, Sci Rep 1 (2020) 1–9. URL: https://doi.org/10.1038/s41598-020-77258-w. doi:10.1038/s41598-020-77258-w.

[12] P. Giannaris, Z. Al-Taie, M. Kovalenko, N. Thanintorn, O. Kholod, Y. Innokenteva, E. Coberly, S. Frazier, K. Laziuk, M. Popescu, C. R. Shyu, D. Xu, R. Hammer, D. Shin, Artificial Intelligence-Driven Structurization of Diagnostic Information in Free-Text Pathology Reports, Journal of Pathology Informatics 11 (2020) 10. URL: https://doi.org/10.4103/jpi.jpi_30_19. doi:10.4103/jpi.jpi_30_19.

[13] J. R. Gregg, M. Lang, L. L. Wang, M. J. Resnick, S. K. Jain, J. L. Warner, D. A. Barocas, Automating the Determination of Prostate Cancer Risk Strata From Electronic Medical Records, JCO Clinical Cancer Informatics 1 (2017) 1–8. URL: https://doi.org/10.1200/CCI.16.00045. doi:10.1200/CCI.16.00045.

[14] A. P. Glaser, B. J. Jordan, J. Cohen, A. Desai, P. Silberman, J. J. Meeks, Automated Extraction of Grade, Stage, and Quality Information From Transurethral Resection of Bladder Tumor Pathology Reports Using Natural Language Processing, JCO Clinical Cancer Informatics 1 (2018) 1–8. URL: https://doi.org/10.1200/CCI.17.00128. doi:10.1200/CCI.17.00128.

[15] K. Roberts, D. Demner-Fushman, E. M. Voorhees, W. R. Hersh, S. Bedrick, A. J. Lazar, S. Pant, Benchmarking Information Retrieval for Precision Oncology: the TREC Precision Medicine Track, in: AMIA 2018, American Medical Informatics Association Annual Symposium, San Francisco, CA, November 3-7, 2018, AMIA, 2018. URL: https://knowledge.amia.org/67852-amia-1.4259402/t006-1.4263223/t006-1.4263224/2976780-1.4263306/2970178-1.4263303.

[16] F. Giachelle, O. Irrera, G. Silvello, MedTAG: a portable and customizable annotation tool for biomedical documents, BMC Medical Informatics Decis. Mak. 21 (2021) 352. URL: https://doi.org/10.1186/s12911-021-01706-4. doi:10.1186/s12911-021-01706-4.

[17] N. Marini, S. Marchesin, S. Otálora, M. Wodzinski, A. Caputo, M. van Rijthoven, W. Aswolinskiy, J. M. Bokhorst, D. Podareanu, E. Petters, S. Boytcheva, G. Buttafuoco, S. Vatrano, F. Fraggetta, J. der Laak, M. Agosti, F. Ciompi, G. Silvello, H. Muller, M. Atzori, Unleashing the potential of digital pathology data by training computer-aided diagnosis models without human annotations, npj Digital Medicine 5 (2022). URL: http://dx.doi.org/10.1038/s41746-022-00635-4. doi:10.1038/s41746-022-00635-4.

[18] L. Chiticariu, Y. Li, F. R. Reiss, Rule-Based Information Extraction is Dead! Long Live Rule-Based Information Extraction Systems!, in: Proc. of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, ACL, 2013, pp. 827–832. URL: https://aclanthology.org/

D13-1079/.

[19] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed Representations of Words and Phrases and their Compositionality, in: Proc. of the 27th Annual Conference on Neural Information Processing Systems 2013, NIPS, Lake Tahoe, Nevada, United States, December 5-8, 2013, 2013, pp. 3111–3119. URL: https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html.

[20] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, Enriching Word Vectors with Subword Information, Trans. Assoc. Comput. Linguistics 5 (2017) 135–146. URL: https://doi.org/10.1162/tacl_a_00051. doi:10.1162/tacl_a_00051.

[21] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep Contextualized Word Representations, in: Proc. of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, ACL, 2018, pp. 2227–2237. URL: https://doi.org/10.18653/v1/n18-1202. doi:10.18653/v1/n18-1202.

[22] J. Devlin, M. W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, CoRR abs/1810.04805 (2018). URL: http://arxiv.org/abs/1810.04805.

[23] G. Campanella, M. G. Hanna, L. Geneslaw, A. Miraflor, V. W. K. Silva, K. J. Busam, E. Brogi, V. E. Reuter, D. S. Klimstra, T. J. Fuchs, Clinical-Grade Computational Pathology using Weakly Supervised Deep Learning on Whole Slide Images, Nat Med 25 (2019) 1301–1309. URL: https://doi.org/10.1038/s41591-019-0508-1. doi:10.1038/s41591-019-0508-1.

[24] M. A. Carbonneau, V. Cheplygina, E. Granger, G. Gagnon, Multiple Instance Learning: A Survey of Problem Characteristics and Applications, Pattern Recognit. 77 (2018) 329–353. URL: https://doi.org/10.1016/j.patcog.2017.10.009. doi:10.1016/j.patcog.2017.10.009.

[25] S. Marchesin, G. Silvello, TBGA: a large-scale gene-disease association dataset for biomedical relation extraction, BMC Bioinform. 23 (2022) 111. URL: https://doi.org/10.1186/s12859-022-04646-6. doi:10.1186/s12859-022-04646-6.

[26] S. Marchesin, Case-Based Retrieval Using Document-Level Semantic Networks, in: Proc. of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018, ACM, 2018, p. 1451. URL: https://doi.org/10.1145/3209978.3210221. doi:10.1145/3209978.3210221.

[27] M. Neumann, D. King, I. Beltagy, W. Ammar, ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing, in: Proc. of the 18th BioNLP Workshop and Shared Task, BioNLP@ACL 2019, Florence, Italy, August 1, 2019, ACL, 2019, pp. 319–327. URL: https://doi.org/10.18653/v1/w19-5034. doi:10.18653/v1/w19-5034.

[28] S. Pyysalo, F. Ginter, H. Moen, T. Salakoski, S. Ananiadou, Distributional Semantics Resources for Biomedical Text Processing, Proc. of LBM (2013) 39–44. URL: https://bio.nlplab.org/pdf/pyysalo13literature.pdf.

[29] W. W. Chapman, W. Bridewell, P. Hanbury, G. F. Cooper, B. G. Buchanan, A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries, J. Biomed. Informatics 34 (2001) 301–310. URL: https://doi.org/10.1006/jbin.2001.1029. doi:10.1006/jbin.2001.1029.

[30] J. W. Ratcliff, D. E. Metzener, Pattern Matching: the Gestalt Approach,

Dr Dobbs Journal 13 (1988) 46. URL: https://www.drdobbs.com/database/pattern-matching-the-gestalt-approach/184407970.

[31] M. Agosti, S. Marchesin, G. Silvello, Learning Unsupervised Knowledge-Enhanced Representations to Reduce the Semantic Gap in Information Retrieval, ACM Trans. Inf. Syst. 38 (2020) 38:1–38:48. URL: https://doi.org/10.1145/3417996. doi:10.1145/3417996.

[32] W. C. Chang, H. F. Yu, K. Zhong, Y. Yang, I. S. Dhillon, Taming pretrained transformers for extreme multi-label text classification, in: KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020, ACM, 2020, pp. 3163–3171. URL: https://doi.org/10.1145/3394486.3403368. doi:10.1145/3394486.3403368.

[33] P. Ruas, V. D. T. Andrade, F. M. Couto, Lasige-biotm at MESINESP2: entity linking with semantic similarity and extreme multi-label classification on spanish biomedical documents, in: Proc. of the Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, Bucharest, Romania, September 21st - to - 24th, 2021, volume 2936 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 324–334. URL: http://ceur-ws.org/Vol-2936/paper-24.pdf.

[34] Y. Zhang, Q. Chen, Z. Yang, H. Lin, Z. Lu, BioWordVec, Improving Biomedical Word Embeddings with Subword Information and MeSH, Scientific Data 6 (2019) 1–9. URL: https://doi.org/10.1038/s41597-019-0055-0. doi:10.1038/s41597-019-0055-0.

[35] E. Alsentzer, J. R. Murphy, W. Boag, W. H. Weng, D. Jin, T. Naumann, M. B. A. McDermott, Publicly Available Clinical BERT Embeddings, CoRR abs/1904.03323 (2019). URL: http://arxiv.org/abs/1904.03323.