

REFERENCES

- [1] Titipat Achakulvisut, Daniel Acuna, and Konrad Kording. 2020. Pubmed parser: a python parser for pubmed open-access xml subset and medline xml dataset xml dataset. *Journal of Open Source Software*, 5, 46, 1979.
- [2] Shabbir Ahmed and Farzana Mithun. 2004. Word stemming to enhance spam filtering. In *CEAS*.
- [3] Rabah Alzaidy, Cornelia Caragea, and C. Lee Giles. 2019. Bi-Istm-crf sequence labeling for keyphrase extraction from scholarly documents. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, 2551–2557.
- [4] M A Andrade and A Valencia. 1998. Automatic extraction of keywords from scientific text: application to the knowledge domain of protein families. *Bioinformatics*, 14, 7, 600–607.
- [5] Samhaa R. El-Beltagy and Ahmed Rafea. 2010. KP-miner: participation in SemEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, 190–193.
- [6] D Blake. 1994. Indexing the british heart journal: choice of keywords. *British heart journal*, 71, 3, 212.
- [7] Seong-Yong Bong and Kyu-Baek Hwang. 2011. Keyphrase extraction in biomedical publications using mesh and intraphrase word co-occurrence information. In *Proceedings of the ACM fifth international workshop on Data and text mining in biomedical informatics*, 63–66.
- [8] Florian Boudin. 2016. Pke: an open source python-based keyphrase extraction toolkit. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, 69–73.
- [9] Florian Boudin. 2018. Unsupervised keyphrase extraction with multipartite graphs. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 667–672.
- [10] Adrien Bougouin, Florian Boudin, and Béatrice Daille. 2013. TopicRank: graph-based topic ranking for keyphrase extraction. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, 543–551.
- [11] Ricardo Campos, Vitor Mangaravite, Arian Pasquali, Alipio Jorge, Célia Nunes, and Adam Jatowt. 2020. Yake! keyword extraction from single documents using multiple local features. *Information Sciences*, 509, 257–289.
- [12] Kathi Canese and Sarah Weis. 2013. Pubmed: the bibliographic database. *The NCBI Handbook*, 2, 1.
- [13] Cornelia Caragea, Florin Adrian Bulgarov, Andreea Godea, and Sujatha Das Gollapalli. 2014. Citation-enhanced keyphrase extraction from research papers: a supervised approach. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1435–1446.
- [14] Jun Chen, Xiaoming Zhang, Yu Wu, Zhao Yan, and Zhoujun Li. 2018. Keyphrase generation with correlation constraints. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 4057–4066.
- [15] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1724–1734.
- [16] Corina Florescu and Cornelia Caragea. 2017. PositionRank: an unsupervised approach to keyphrase extraction from scholarly documents. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1105–1115.
- [17] Ygor Gallina, Florian Boudin, and Béatrice Daille. 2020. Large-scale evaluation of keyphrase extraction models. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, 271–278.
- [18] Zelalem Gero and Joyce Ho. 2021. Word centrality constrained representation for keyphrase extraction. In *Proceedings of the 20th Workshop on Biomedical Language Processing*, 155–161.
- [19] Zelalem Gero and Joyce C Ho. 2021. Uncertainty-based self-training for biomedical keyphrase extraction. In *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, 1–4.
- [20] Zelalem Gero and Joyce C. Ho. 2019. Namedkeys: unsupervised keyphrase extraction for biomedical documents. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (ICB '19)*. Niagara Falls, NY, USA, 328–337.
- [21] Isidoro Gil-Leiva and Adolfo Alonso-Arroyo. 2007. Keywords given by authors of scientific articles in database descriptors. *Journal of the American society for information science and technology*, 58, 8, 1175–1187.
- [22] Anette Hulth. 2003. Improved automatic keyword extraction given more linguistic knowledge. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, 216–223.
- [23] Zhenchao Jiang, Lishuang Li, Degen Huang, and Liuke Jin. 2015. Training word embeddings for deep learning in biomedical text mining tasks. In *2015 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE, 625–628.
- [24] Sun Kim, Lana Yeganova, Donald C Comeau, W John Wilbur, and Zhiyong Lu. 2018. Pubmed phrases, an open set of coherent phrases for searching biomedical literature. *Scientific data*, 5, 1, 1–11.
- [25] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36, 4, 1234–1240.
- [26] Edward Loper and Steven Bird. 2002. NLTK: the natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, 63–70.
- [27] Vanessa Loza, Shibamouli Lahiri, Rada Mihalcea, and Po-Hsiang Lai. 2014. Building a dataset for summarization and keyword extraction from emails. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 2441–2446.
- [28] Rui Meng, Sanqiang Zhao, Shuguang Han, Daqing He, Peter Brusilovsky, and Yu Chi. 2017. Deep keyphrase generation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 582–592.
- [29] Rui Meng, Sanqiang Zhao, Shuguang Han, Daqing He, Peter Brusilovsky, and Yu Chi. 2017. Deep keyphrase generation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 582–592.
- [30] Rada Mihalcea and Paul Tarau. 2004. TextRank: bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 404–411.
- [31] Justin Samuel Payan. 2018. *Keyphrase extraction from scientific literature using joint geometric graph embedding matching*. Ph.D. Dissertation. University of Georgia.
- [32] Jakub Piskorski, Nicolas Stefanovitch, Guillaume Jacquet, and Aldo Podavini. 2021. Exploring linguistically-lightweight keyword extraction techniques for indexing news articles in a multilingual set-up. In *Proceedings of the EACL Hackshop on News Media Content Analysis and Automated Report Generation*, 35–44.
- [33] Martin F. Porter. 2001. Snowball: a language for stemming algorithms. (2001).
- [34] Juan Ramos et al. 2003. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning number 1*. Vol. 242, 29–48.
- [35] Dhruva Sahrawat et al. 2020. Keyphrase extraction as sequence labeling using contextualized embeddings. *Advances in Information Retrieval*, 12036, 328.
- [36] Tokala Yaswanth Sri Sai Santosh, Debarshi Kumar Sanyal, Plaban Kumar Bhowmick, and Partha Pratim Das. 2020. Dake: document-level attention for keyphrase extraction. *Advances in Information Retrieval*, 12036, 392.
- [37] Kamal Sarkar. 2013. A hybrid approach to extract keyphrases from medical documents. *International Journal of Computer Applications*, 63, 18, 14–19.
- [38] Alexander Thorsten Schutz et al. 2008. Keyphrase extraction from single documents in the open domain exploiting linguistic and statistical methods. *M. App. Sc Thesis*.
- [39] Jasmeet Singh and Vishal Gupta. 2016. Text stemming: approaches, applications, and challenges. *ACM Comput. Surv.*, 49, 3, Article 45, 46 pages.
- [40] Lucas Sterckx, Thomas Demeester, Johannes Deleu, and Chris Develder. 2015. Topical word importance for fast keyphrase extraction. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion)*. Florence, Italy, 121–122.
- [41] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, 3104–3112.
- [42] Siw Waffenschmidt, Marco Knelangen, Wiebke Sieben, Stefanie Bühn, and Dawid Pieper. 2019. Single screening versus conventional double screening for study selection in systematic reviews: a methodological systematic review. *BMC Medical Research Methodology*, 19, 132.
- [43] Xiaojun Wan and Jianguo Xiao. 2008. CollabRank: towards a collaborative approach to single-document keyphrase extraction. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, 969–976.
- [44] Yue Wang, Jing Li, Hou Pong Chan, Irwin King, Michael R. Lyu, and Shuming Shi. 2019. Topic-aware neural keyphrase generation for social media language. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2516–2526.
- [45] Thomas Wolf et al. 2020. Transformers: state-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45.