

































































- [53] Navigli R, Velardi P. Semantic interpretation of terminological strings. In: Proc. of the 6th Int'l Conf. on Terminology and Knowledge Engineering. 2002. 95–100.
- [54] Liu L, Xiao YY. A statistical domain terminology extraction method based on word length and grammatical feature. Journal of Harbin Engineering University, 2017,38(9):1437–1443 (in Chinese with English abstract).
- [55] Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. Information Processing & Management, 1988,24(5): 513–523. [doi: 10.1016/0306-4573(88)90021-0]
- [56] Zhou L, Shi SM, Feng C, Huang HY. A Chinese term extraction system based on multi-strategies integration. Journal of the China Society for Scientific and Technical Information, 2010,29(3):460–467 (in Chinese with English abstract).
- [57] Yan XL, Liu YQ, Fang Q, Zhang M, Ma SP, Ru LY. Domain-specific terms extraction based on Web resource and user behavior. Ruan Jian Xue Bao/Journal of Software, 2013,24(9):2089–2100 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4358.htm> [doi: 10.3724/SP.J.1001.2013.04358]
- [58] Lossio-Ventura JA, Jonquet C, Roche M, Teisseire M. Biomedical terminology extraction: A new combination of statistical and web mining approaches. In: Proc. of the JADT 2014. 2014. 421–432.
- [59] Church K, Gale W. Inverse document frequency (IDF): A measure of deviations from poisson. In: Natural Language Processing Using Very Large Corpora. Dordrecht: Springer-Verlag, 1999. 283–295. [doi: 10.1007/978-94-017-2390-9\_18]
- [60] Li LS, Dang YZ, Zhang J, Li D. Domain term extraction based on conditional random fields combined with active learning strategy. Journal of Information & Computational Science, 2012,9(7):1931–1940.
- [61] Rose S, Engel D, Cramer N, Cowley W. Automatic keyword extraction from individual documents. In: Text Mining: Applications and Theory, 2010. 1–20.
- [62] Bordea G, Buitelaar P, Polajnar T. Domain-independent term extraction through domain modelling. In: Proc. of the 10th Int'l Conf. on Terminology and Artificial Intelligence. 2013.
- [63] Astrakhantsev N. Methods and software for terminology extraction from domain-specific text collection [Ph.D. Thesis]. Institute for System Programming of Russian Academy of Sciences, 2015.
- [64] You HL, Zhang W, Shen JY, Liu T. A weighted voting based automatic term recognition method. Journal of Chinese Information Processing, 2011,25(3):9–17 (in Chinese with English abstract).
- [65] He L. Domain ontology terminology extraction based on integrated strategy method. Journal of the China Society for Scientific and Technical Information, 2012,31(8):798–804 (in Chinese with English abstract).
- [66] Li LS, Wang YW, Huang DG. Term extraction based on information entropy and word frequency distribution variety. Journal of Chinese Information Processing, 2015,29(1):82–87 (in Chinese with English abstract).
- [67] Stanković R, Krstev C, Obradović I, Lazic B. Rule-based automatic multi-word term extraction and lemmatization. In: Calzolari N, Choukri K, eds. Proc. of the LREC. Portoro: European Language Resources Association, 2016. 507–514.
- [68] Dong YY, Li WH, Hu H. Domain term extraction method based on hierarchical combination strategy for Chinese Web documents. Journal of Northwestern Polytechnical University, 2017,35(4):729–735 (in Chinese with English abstract).
- [69] Pazienza MT, Pennacchiotti M, Zanzotto FM. Terminology extraction: An analysis of linguistic and statistical approaches. In: Knowledge Mining. Berlin: Springer-Verlag, 2005. 255–279. [doi: 10.1007/3-540-32394-5\_20]
- [70] Ahmad K, Gillam L, Tostevin L. University of surrey participation in trec8: Weirdness indexing for logical document extrapolation and retrieval (wilder). In: Proc. of the TREC. 1999. 1–8.
- [71] Peñas A, Verdejo F, Gonzalo J. Corpus-based terminology extraction applied to information access. In: Proc. of the Corpus Linguistics. 2001. 458–465.
- [72] Park Y, Byrd RJ, Boguraev BK. Automatic glossary extraction: Beyond terminology identification. In: Proc. of the COLING. Stroudsburg: ACL, 2002. 1–7. [doi: 10.3115/1072228.1072370]
- [73] Sclano F, Velardi P. Termextractor: A Web application to learn the shared terminology of emergent web communities. In: Proc. of the 3th Int'l Conf. on Interoperability for Enterprise Software and Applications. London: Springer-Verlag, 2007. 287–290.
- [74] Lopes L, Fernandes P, Vieira R. Estimating term domain relevance through term frequency, disjoint corpora frequency-tf-dcf. Knowledge-based Systems, 2016,97:237–249.

- [75] Mykowiecka A, Marciniak M, Rychlik P. Recognition of irrelevant phrases in automatically extracted lists of domain terms. *Int'l Journal of Theoretical and Applied Issues in Specialized Communication*, 2018,24(1):66–90.
- [76] Vivaldi J, Rodríguez H. Using Wikipedia for term extraction in the biomedical domain: First experiences. *Procesamiento del Lenguaje Natural*, 2010,45:251–254.
- [77] Haque R, Penkale S, Way A. TermFinder: Log-likelihood comparison and phrase-based statistical machine translation models for bilingual terminology extraction. *Language Resources and Evaluation*, 2018,52(2):365–400. [doi: 10.1007/s10579-018-9412-4]
- [78] Zheng D, Zhao T, Yang J. Research on domain term extraction based on conditional random fields. In: *Proc. of the ICCPOL*. Berlin: Springer-Verlag, 2009. 290–296.
- [79] Zhang X, Song Y, Fang AC. Term recognition using conditional random fields. In: *Proc. of the 6th Int'l Conf. on Natural Language Processing and Knowledge Engineering*. IEEE, 2010. 1–6.
- [80] Zhang ZC. Using integration strategy and multi-level termhood to extract terminology. *Journal of the China Society for Scientific and Technical Information*, 2011,28(3):275–285 (in Chinese with English abstract).
- [81] Loukachevitch NV. Automatic term recognition needs multiple evidence. In: Calzolari N, Choukri K, eds. *Proc. of the LREC*. Portoro: European Language Resources Association, 2012. 2401–2407.
- [82] Conrado MD, Pardo TA, Rezende SO. A machine learning approach to automatic term extraction using a rich feature set. In: *Proc. of the 2013 NAACL HLT Student Research Workshop*. Stroudsburg: ACL, 2013. 16–23.
- [83] Yuan Y, Gao J, Zhang Y. Supervised learning for robust term extraction. In: *Proc. of the Int'l Conf. on Asian Language Processing*. IEEE, 2017. 302–305.
- [84] Yang Y, Yu H, Meng Y, Lu Y, Xia Y. Fault-tolerant learning for term extraction. In: *Proc. of the 24th Pacific Asia Conf. on Language, Information and Computation*. Institute for Digital Enhancement of Cognitive Development, 2010. 321–330.
- [85] Maldonado A, Lewis D. Self-tuning ongoing terminology extraction retrained on terminology validation decisions. In: *Proc. of the Conf. on Terminology and Knowledge Engineering*. 2016. 91–101.
- [86] Aker A, Paramita M, Gaizauskas R. Extracting bilingual terminologies from comparable corpora. In: *Proc. of the 51st Annual Meeting of the Association for Computational Linguistics*. Stroudsburg: ACL, 2013. 402–411.
- [87] Wang H, Wang MP, Su XN. A study on Chinese patent terms extraction for ontology learning. *Journal of the China Society for Scientific and Technical Information*, 2016,35(6):573–585 (in Chinese with English abstract).
- [88] Khosla K, Jones R, Bowman N. Featureless deep learning methods for automated key-term extraction. Stanford: Stanford University, 2019. 1–10.
- [89] Gao Y, Yuan Y. Feature-less end-to-end nested term extraction. In: *Proc. of the CCF Int'l Conf. on Natural Language Processing and Chinese Computing*. Cham: Springer-Verlag, 2019. 607–616.
- [90] Zhao H, Wang F. A deep learning model and self-training algorithm for theoretical terms extraction. *Journal of the China Society for Scientific and Technical Information*, 2018,37(9):923–938 (in Chinese with English abstract).
- [91] Kucza M, Niehues J, Zenkel T, Waibel A, Stüker S. Term extraction via neural sequence labeling a comparative evaluation of strategies using recurrent neural networks. In: *Proc. of the Interspeech*. Hyderabad: ISCA, 2018. 2072–2076.
- [92] Lossio-Ventura JA, Jonquet C, Roche M, Teisseire M. Biomedical term extraction: Overview and a new methodology. *Information Retrieval Journal*, 2016,19(1-2):59–99. [doi: 10.1007/s10791-015-9262-2]
- [93] Bikel D, Zitouni I. *Multilingual Natural Language Processing Applications: From Theory to Practice*. IBM Press, 2012.
- [94] Yang K, Ding X, Zhang Y, Chen L, Zheng B, Gao Y. Distributed similarity queries in metric spaces. *Data Science and Engineering*, 2019,4(2):93–108.
- [95] El-Beltagy SR, Rafea A. KP-miner: Participation in semeval-2. In: *Proc. of the 5th Int'l Workshop on Semantic Evaluation*. Stroudsburg: ACL, 2010. 190–193.
- [96] Yu Y, Zhao NX. Patent term extraction based on generic words and term components. *Journal of the China Society for Scientific and Technical Information*, 2018,37(7):742–752 (in Chinese with English abstract).
- [97] Lahbib W, Bounhas I, Slimani Y. A possibilistic approach for Arabic domain terminology extraction and translation. In: *Proc. of the Int'l Symp. on Computer and Information Sciences*. Cham: Springer-Verlag, 2018. 231–238.

- [98] Li K, Zha H, Su Y, Yan X. Concept mining via embedding. In: Proc. of the 2018 IEEE Int'l Conf. on Data Mining. Singapore: IEEE Computer Society, 2018. 267–276.
- [99] Khan MT, Ma Y, Kim J. Term ranker: A graph-based re-ranking approach. In: Proc. of the 29th Int'l Florida Artificial Intelligence Research Society Conf. Florida: AAAI Press, 2016. 310–315.
- [100] Conde A, Larrañaga M, Arruarte A, Elorriaga JA, Roth D. LiTeWi: A combined term extraction and entity linking method for eliciting educational ontologies from textbooks. *Journal of the Association for Information Science and Technology*, 2016,67(2): 380–399.
- [101] Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: Bringing order to the Web. Stanford InfoLab, 1999. <http://courses.washington.edu/ir2010/readings/page.pdf>
- [102] Pan LM, Wang XC, Li JZ, Tang J. Course concept extraction in MOOCs via embedding-based graph propagation. In: Proc. of the 8th Int'l Joint Conf. on Natural Language Processing. Asian Federation of Natural Language Processing, 2017. 875–884.
- [103] Zhang Z, Gao J, Ciravegna F. Semre-rank: Improving automatic term extraction by incorporating semantic relatedness with personalised pagerank. *ACM Trans. on Knowledge Discovery from Data*, 2018,12(5):1–41. [doi: 10.1145/3201408]
- [104] Zhang Z, Petrak J, Maynard D. Adapted textrank for term extraction: A generic method of improving automatic term extraction algorithms. In: Proc. of the 14th Int'l Conf. on Semantic Systems. Elsevier, 2018. 102–108.
- [105] Su MS, Li L, Liu ZY. Unsupervised bilingual terminology extraction algorithm for Chinese-English parallel patents. *Journal of Tsinghua University (Science and Technology)*, 2014,54(10):1339–1343 (in Chinese with English abstract).
- [106] Li B, Wang B, Zhou R, Yang X, Liu C. CITPM: A cluster-based iterative topical phrase mining framework. In: Proc. of the Int'l Conf. on Database Systems for Advanced Applications. Switzerland: Springer-Verlag, 2016. 197–213.
- [107] Arora C, Sabetzadeh M, Briand L, Zimmer F. Automated extraction and clustering of requirements glossary terms. *IEEE Trans. on Software Engineering*, 2017,43(10):918–945.
- [108] Kim JD, Ohta T, Tateisi Y, Tsujii J. GENIA corpus—A semantically annotated corpus for bio-textmining. In: Proc. of the 11th Int'l Conf. on Intelligent Systems for Molecular Biology. 2003. 180–182.
- [109] Medelyan O, Witten IH. Domain-independent automatic keyphrase indexing with small training sets. *Journal of the American Society for Information Science and Technology*, 2008,59(7):1026–1040. [doi: 10.1002/asi.20790]
- [110] Krapivin M, Autaeu A, Marchese M. Large dataset for keyphrases extraction. DISI-09-055, DISI, University of Trento, 2009.
- [111] Handschuh S, QasemiZadeh B. The ACL RD-TEC: A dataset for benchmarking terminology extraction and classification in computational linguistics. In: Proc. of the 4th Int'l Workshop on Computational Terminology. Stroudsburg: ACL, 2014. 52–63.
- [112] QasemiZadeh B, Schumann AK. The ACL RD-TEC 2.0: A language resource for evaluating term extraction and entity recognition methods. In: Calzolari N, Choukri K, eds. Proc. of the LREC. Portoro: European Language Resources Association, 2016. 1862–1868.
- [113] Blancafort H, Daille B, Gornostay T, Heid U, Sharoff S, Méchoulam C. TTC: Terminology extraction, translation tools and comparable corpora. In: Proc. of the 14th EuraLex Int'l Congress. 2010. 263–268.
- [114] Koehn P. Europarl: A parallel corpus for statistical machine translation. *MT Summit*, 2005,5:79–86. <http://www.statmt.org/europarl/>
- [115] Zhang Z, Iria J, Brewster C, Ciravegna F. A comparative evaluation of term recognition algorithms. In: Calzolari N, Choukri K, eds. Proc. of the LREC. Portoro: European Language Resources Association, 2008. 28–30.
- [116] Cram D, Daille B. TermSuit: Terminology extraction with term variant detection. In: Proc. of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2016. 13–18. [doi: 10.18653/v1/P16-4003]
- [117] Oliver A, Vázquez M. TBXTools: A free, fast and flexible tool for automatic terminology extraction. In: Proc. of the Int'l Conf. Recent Advances in Natural Language Processing. 2015. 473–479.
- [118] Lossio-Ventura JA, Jonquet C, Roche M, Teisseire M. BIOTEX: A system for biomedical terminology extraction, ranking, and validation. In: Proc. of the 13th Int'l Semantic Web Conference. CEUR-WS.org, 2014. 157–160.
- [119] Spasić I, Greenwood M, Preece A, Francis N, Elwyn G. FlexiTerm: A flexible term recognition method. *Journal of Biomedical Semantics*, 2013,4(1):27–43. [doi: 10.1186/2041-1480-4-27]



- [120] Verberne S, Sappelli M, Hiemstra D, Kraaij W. Evaluation and analysis of term scoring methods for term extraction. *Information Retrieval Journal*, 2016,19(5):510–545. [doi: 10.1007/s10791-016-9286-2]

#### 附中文参考文献:

- [30] 袁劲松,张小明,李舟军.术语自动抽取方法研究综述. *计算机科学*,2015,42(8):7–12.
- [40] 李思良,许斌,杨玉基.DRTE:面向基础教育的术语抽取方法. *中文信息学报*,2018,32(3):101–109.
- [54] 刘里,肖迎元.基于术语长度和语法特征的统计领域术语抽取. *哈尔滨工程大学学报*,2017,38(9):1437–1443.
- [56] 周浪,史树敏,冯冲,黄河燕.基于多策略融合的中文术语抽取方法. *情报学报*,2010,29(3):460–467.
- [57] 闫兴龙,刘奕群,方奇,张敏,马少平,茹立云.基于网络资源与用户行为信息的领域术语提取. *软件学报*,2013,24(9):2089–2100. <http://www.jos.org.cn/1000-9825/4358.htm> [doi: 10.3724/SP.J.1001.2013.04358]
- [64] 游宏梁,张巍,沈钧毅,刘挺.一种基于加权投票的术语自动识别方法. *中文信息学报*,2011,25(3):9–17.
- [65] 何琳.基于多策略的领域本体术语抽取研究. *情报学报*,2012,31(8):798–804.
- [66] 李丽双,王意文,黄德根.基于信息熵和词频分布变化的术语抽取研究. *中文信息学报*,2015,29(1):82–87.
- [68] 董洋溢,李伟华,于会.文本特征和复合统计量的领域术语抽取方法. *西北工业大学学报*,2017,35(4):729–735.
- [80] 章成志.基于多层次语义度的一体化术语抽取研究. *情报学报*,2011,28(3):275–285.
- [87] 王昊,王密平,苏新宁.面向本体学习的中文专利术语抽取研究. *情报学报*,2016,35(6):573–585.
- [90] 赵洪,王芳.理论术语抽取的深度学习模型及自训练算法研究. *情报学报*,2018,37(9):923–938.
- [96] 俞琰,赵乃瑄.基于通用词与术语部件的专利术语抽取. *情报学报*,2018,37(7):742–752.
- [105] 孙茂松,李莉,刘知远.面向中英平行专利的双语术语自动抽取. *清华大学学报:自然科学版*,2014,54(10):1339–1343.



张雪(1989—),女,博士,主要研究领域为自然语言处理,数据挖掘.



李翠平(1971—),女,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为社会网络分析,社会推荐,大数据分析和挖掘.



孙宏宇(1994—),男,硕士,主要研究领域为数据挖掘,自然语言处理.



陈红(1965—),女,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为大数据管理与隐私保护,基于新硬件的数据管理与数据分析,数据仓库与数据挖掘.



辛东兴(1994—),男,硕士,主要研究领域为数据挖掘,自然语言处理.