Full text available at: http://dx.doi.org/10.1561/1900000064

Machine Knowledge: Creation and Curation of Comprehensive Knowledge Bases

Full text available at: http://dx.doi.org/10.1561/1900000064

Other titles in Foundations and Trends® in Databases

Cloud Data Services: Workloads, Architectures and Multi-Tenancy
Vivek Narasayya and Surajit Chaudhuri
ISBN: 978-1-68083-774-2

Data Provenance
Boris Glavic
ISBN: 978-1-68083-828-2

FPGA-Accelerated Analytics: From Single Nodes to Clusters
Zsolt István, Kaan Kara and David Sidler
ISBN: 978-1-68083-734-6

Distributed Learning Systems with First-Order Methods
Ji Liu and Ce Zhango
ISBN: 978-1-68083-700-1

Full text available at: http://dx.doi.org/10.1561/1900000064

Machine Knowledge: Creation and Curation of Comprehensive Knowledge Bases

Gerhard Weikum
Max Planck Institute for Informatics
weikum@mpi-inf.mpg.de

Xin Luna Dong
Amazon
lunadong@amazon.com

Simon Razniewski
Max Planck Institute for Informatics
srazniew@mpi-inf.mpg.de

Fabian Suchanek
Telecom Paris University
suchanek@telecom-paris.fr

now
the essence of knowledge
Boston — Delft

Full text available at: http://dx.doi.org/10.1561/1900000064

Foundations and Trends® in Databases

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

G. Weikum and L. Dong and S. Razniewski and F. Suchanek. *Machine Knowledge: Creation and Curation of Comprehensive Knowledge Bases*. Foundations and Trends® in Databases, vol. 10, no. 2-4, pp. 108–490, 2021.

ISBN: 978-1-68083-837-4
© 2021 G. Weikum and L. Dong and S. Razniewski and F. Suchanek

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1 781 871 0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Full text available at: http://dx.doi.org/10.1561/1900000064

References

[1] D. B. Lenat and E. A. Feigenbaum, "On the thresholds of knowledge", *Artificial Intelligence (AI)*, vol. 47, no. 1-3, 1991, pp. 185–250. DOI: 10.1016/0004-3702(91)90055-O.

[2] D. B. Lenat, "CYC: A large-scale investment in knowledge infrastructure", *Communication of the ACM*, vol. 38, no. 11, 1995, pp. 32–38. DOI: 10.1145/219717.219745.

[3] C. Fellbaum and G. A. Miller, *WordNet: An electronic lexical database*. MIT press, 1998.

[4] S. Staab and R. Studer, *Handbook on Ontologies*. Springer, 2009. DOI: 10.1007/978-3-540-92673-3.

[5] R. V. Guha, D. Brickley, and S. Macbeth, "Schema.org: Evolution of structured data on the web", *Communication of the ACM*, vol. 59, no. 2, 2016, pp. 44–51. DOI: 10.1145/2844544.

[6] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. G. Ives, "Dbpedia: A nucleus for a web of open data", in *International Semantic Web Conference (ISWC)*, 2007. DOI: 10.1007/978-3-540-76298-0_52.

[7] K. D. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: A collaboratively created graph database for structuring human knowledge", in *ACM Conference on Management of Data (SIGMOD)*, 2008. DOI: 10.1145/1376616.1376746.

Full text available at: http://dx.doi.org/10.1561/1900000064

[8] O. Etzioni, M. J. Cafarella, D. Downey, A. Popescu, T. Shaked, S. Soderland, D. S. Weld, and A. Yates, "Unsupervised named-entity extraction from the web: An experimental study", *Artificial Intelligence (AI)*, vol. 165, no. 1, 2005, pp. 91–134. DOI: 10.1016/j.artint.2005.03.001.

[9] N. N. Dalvi, R. Kumar, B. Pang, R. Ramakrishnan, A. Tomkins, P. Bohannon, S. S. Keerthi, and S. Merugu, "A web of concepts", in *ACM Symposium on Principles of Database Systems (PODS)*, 2009. DOI: 10.1145/1559795.1559797.

[10] S. P. Ponzetto and M. Strube, "Deriving a large-scale taxonomy from wikipedia", in *Conference on Artificial Intelligence (AAAI)*, 2007. [Online]. Available: http://www.aaai.org/Library/AAAI/2007/aaai07-228.php.

[11] F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: A core of semantic knowledge", in *The Web Conference (WWW)*, 2007. DOI: 10.1145/1242572.1242667.

[12] R. Navigli and S. P. Ponzetto, "Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network", *Artificial Intelligence (AI)*, vol. 193, 2012, pp. 217–250. DOI: 10.1016/j.artint.2012.07.001.

[13] R. Speer and C. Havasi, "Representing general relational knowledge in conceptnet 5", in *Conference on Language Resources and Evaluation (LREC)*, 2012. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2012/summaries/1072.html.

[14] J. Shin, S. Wu, F. Wang, C. D. Sa, C. Zhang, and C. Ré, "Incremental knowledge base construction using deepdive", *Proceedings of the VLDB Endowment*, vol. 8, no. 11, 2015, pp. 1310–1321. DOI: 10.14778/2809974.2809991.

[15] Z. Nie, J. Wen, and W. Ma, "Statistical entity extraction from the web", *Proceedings of the IEEE*, vol. 100, no. 9, 2012, pp. 2675–2687. DOI: 10.1109/JPROC.2012.2191369.

[16] X. L. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun, and W. Zhang, "Knowledge vault: A web-scale approach to probabilistic knowledge fusion", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2014. DOI: 10.1145/2623330.2623623.

Full text available at: http://dx.doi.org/10.1561/1900000064

[17] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. H. Jr., and T. M. Mitchell, "Toward an architecture for never-ending language learning", in *Conference on Artificial Intelligence (AAAI)*, 2010. [Online]. Available: http://www.aaai.org/ocs/index.php/AAAI/AAAI10/paper/view/1879.

[18] W. Wu, H. Li, H. Wang, and K. Q. Zhu, "Probase: A probabilistic taxonomy for text understanding", in *ACM Conference on Management of Data (SIGMOD)*, 2012. DOI: 10.1145/2213836.2213891.

[19] S. Hertling and H. Paulheim, "Webisalod: Providing hypernymy relations extracted from the web as linked open data", in *International Semantic Web Conference (ISWC)*, 2017. DOI: 10.1007/978-3-319-68204-4_11.

[20] D. Vrandecic and M. Krötzsch, "Wikidata: A free collaborative knowledgebase", *Communication of the ACM*, vol. 57, no. 10, 2014, pp. 78–85. DOI: 10.1145/2629489.

[21] Z. Wang, J. Li, Z. Wang, S. Li, M. Li, D. Zhang, Y. Shi, Y. Liu, P. Zhang, and J. Tang, "Xlore: A large-scale english-chinese bilingual knowledge graph", in *International Semantic Web Conference (ISWC)*, 2013. [Online]. Available: http://ceur-ws.org/Vol-1035/iswc2013%5C_demo%5C_31.pdf.

[22] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutierrez, J. E. L. Gayo, S. Kirrane, S. Neumaier, A. Polleres, R. Navigli, A. N. Ngomo, S. M. Rashid, A. Rula, L. Schmelzeisen, J. F. Sequeda, S. Staab, and A. Zimmermann, "Knowledge graphs", *arXiv*, vol. abs/2003.02320, 2020.

[23] T. Heath and C. Bizer, *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool Publishers, 2011. DOI: 10.2200/S00334ED1V01Y201102WBE001.

[24] A. Hogan, *The Web of Data*. Springer, 2020. DOI: 10.1007/978-3-030-51580-5.

[25] N. Noy, Y. Gao, A. Jain, A. Narayanan, A. Patterson, and J. Taylor, "Industry-scale knowledge graphs: Lessons and challenges", *Communication of the ACM*, vol. 62, no. 8, 2019, pp. 36–43.

Full text available at: http://dx.doi.org/10.1561/1900000064

[26] H. Bast, B. Buchhold, and E. Haussmann, "Semantic search on text and knowledge bases", *Foundations and Trends in Information Retrieval*, vol. 10, no. 2-3, 2016, pp. 119–271. DOI: 10.1561/1500000032.

[27] R. Reinanda, E. Meij, and M. de Rijke, "Knowledge graphs: An information retrieval perspective", *Foundations and Trends in Information Retrieval*, 2020.

[28] Q. Guo, F. Zhuang, C. Qin, H. Zhu, X. Xie, H. Xiong, and Q. He, "A survey on knowledge graph-based recommender systems", *CoRR*, vol. abs/2003.00911, 2020.

[29] A. Singhal, "Introducing the knowledge graph: Things, not strings", in *Google Blog, 16 May 2012, https://www.blog.google/products/search/introducing-knowledge-graph-things-not/*.

[30] D. A. Ferrucci *et al.*, "Special issue on "this is watson"", *IBM Journal of Research and Development*, vol. 56, no. 3, 2012, p. 1. DOI: 10.1147/JRD.2012.2184356.

[31] X. L. Dong, "Building a broad knowledge graph for products", in *Slides from Keynote at International Conference on Data Engineering (ICDE), http://lunadong.com/talks/BG.pdf*, 2019.

[32] X. L. Dong, X. He, A. Kan, X. Li, Y. Liang, J. Ma, Y. E. Xu, C. Zhang, T. Zhao, G. B. Saldana, S. Deshpande, A. M. Manduca, J. Ren, S. P. Singh, F. Xiao, H. Chang, G. Karamanolakis, Y. Mao, Y. Wang, C. Faloutsos, A. McCallum, and J. Han, "Autoknow: Self-driving knowledge collection for products of thousands of types", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2020. [Online]. Available: https://dl.acm.org/doi/10.1145/3394486.3403323.

[33] X. Luo, L. Liu, Y. Yang, L. Bo, Y. Cao, J. Wu, Q. Li, K. Yang, and K. Q. Zhu, "Alicoco: Alibaba e-commerce cognitive concept net", in *ACM Conference on Management of Data (SIGMOD)*, ACM, 2020. DOI: 10.1145/3318464.3386132.

Full text available at: http://dx.doi.org/10.1561/1900000064

[34] Baidu, "Introducing qian yan, baidu's new plan to build 100 chinese nlp datasets in three years", in *Baidu Research Blog, 28 August 2020, http://research.baidu.com/Blog/index-view?id=146*.

[35] R. Qiang, "Understand your world with bing", in *Bing Blog, 21 March 2013, https://blogs.bing.com/search/2013/03/21/understand-your-world-with-bing/*.

[36] M. B. Hoy, "Wolfram|alpha: A brief introduction", *Medical reference services quarterly*, vol. 29, no. 1, 2010, pp. 67–74.

[37] E. Meij, "Understanding news using the bloomberg knowledge graph", in *Talk at Big Data Innovators Gathering (BIG) at the Web Conference 2019, https://speakerdeck.com/emeij/understanding-news-using-the-bloomberg-knowledge-graph*.

[38] A. Doan, A. Y. Halevy, and Z. G. Ives, *Principles of Data Integration*. Morgan Kaufmann, 2012. [Online]. Available: http://research.cs.wisc.edu/dibook/.

[39] A. Doan, P. Konda, P. S. G. C., Y. Govind, D. Paulsen, K. Chandrasekhar, P. Martinkus, and M. Christie, "Magellan: Toward building ecosystems of entity matching solutions", *Commun. ACM*, vol. 63, no. 8, 2020, pp. 83–91. DOI: 10.1145/3405476.

[40] D. Diefenbach, V. López, K. D. Singh, and P. Maret, "Core techniques of question answering systems over knowledge bases: A survey", *Knowledge and Information Systems (KAIS)*, vol. 55, no. 3, 2018, pp. 529–569. DOI: 10.1007/s10115-017-1100-y.

[41] C. Lei, F. özcan, A. Quamar, A. R. Mittal, J. Sen, D. Saha, and K. Sankaranarayanan, "Ontology-based natural language query interfaces for data exploration", *IEEE Data Engineering Bulletin*, vol. 41, no. 3, 2018, pp. 52–63. [Online]. Available: http://sites.computer.org/debull/A18sept/p52.pdf.

[42] C. Unger, A. Freitas, and P. Cimiano, "An introduction to question answering over linked data", in *Reasoning Web Summer School*, Springer, 2014. DOI: 10.1007/978-3-319-10587-1_2.

Full text available at: http://dx.doi.org/10.1561/1900000064

[43] D. Adiwardana, M. Luong, D. R. So, J. Hall, N. Fiedel, R. Thoppilan, Z. Yang, A. Kulshreshtha, G. Nemade, Y. Lu, and Q. V. Le, "Towards a human-like open-domain chatbot", *CoRR*, vol. abs/2001.09977, 2020. [Online]. Available: https://arxiv.org/abs/2001.09977.

[44] W. Shen, J. Wang, and J. Han, "Entity linking with a knowledge base: Issues, techniques, and solutions", *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 27, no. 2, 2015, pp. 443–460. DOI: 10.1109/TKDE.2014.2327028.

[45] F. M. Suchanek and N. Preda, "Semantic culturomics (vision paper)", *Proceedings of the VLDB Endowment*, vol. 7, no. 12, 2014. DOI: 10.14778/2732977.2732994.

[46] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and F. Li, "Imagenet: A large-scale hierarchical image database", in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009. DOI: 10.1109/CVPRW.2009.5206848.

[47] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L. Li, D. A. Shamma, M. S. Bernstein, and L. Fei-Fei, "Visual genome: Connecting language and vision using crowdsourced dense image annotations", *International Journal of Computer Vision (IJCV)*, vol. 123, no. 1, 2017, pp. 32–73. DOI: 10.1007/s11263-016-0981-7.

[48] E. Rahm and H. H. Do, "Data cleaning: Problems and current approaches", *IEEE Data Engineering Bulletin*, vol. 23, no. 4, 2000, pp. 3–13. [Online]. Available: http://sites.computer.org/debull/A00DEC-CD.pdf.

[49] I. F. Ilyas and X. Chu, "Trends in cleaning relational data: Consistency and deduplication", *Foundations and Trends in Databases*, vol. 5, no. 4, 2015, pp. 281–393. DOI: 10.1561/1900000045.

[50] I. F. Ilyas and X. Chu, *Data Cleaning*. ACM, 2019. DOI: 10.1145/3310205.

[51] G. Li, "Special issue on large-scale data integration", *IEEE Data Engineering Bulletin*, vol. 41, no. 2, 2018, p. 2. [Online]. Available: http://sites.computer.org/debull/A18june/p2.pdf.

Full text available at: http://dx.doi.org/10.1561/1900000064

[52] A. Y. Halevy, M. J. Franklin, and D. Maier, "Principles of dataspace systems", in *ACM Symposium on Principles of Database Systems (PODS)*, 2006. DOI: 10.1145/1142351.1142352.

[53] S. J. Russell and P. Norvig, *Artificial intelligence - a modern approach, 2nd Edition*. Prentice Hall, 2003. [Online]. Available: http://www.worldcat.org/oclc/314283679.

[54] F. M. Suchanek, J. Lajus, A. Boschin, and G. Weikum, "Knowledge representation and rule mining in entity-centric knowledge bases", in *Reasoning Web Summer School*, Springer, 2019. DOI: 10.1007/978-3-030-31423-1_4.

[55] W3C, *RDF 1.1 Primer, https://www.w3.org/TR/rdf11-primer/*. World Wide Web Consortium, 2014.

[56] W3C, *RDF Schema, https://www.w3.org/TR/rdf-schema/*. World Wide Web Consortium, 2014.

[57] I. Gurevych, J. Eckle-Kohler, and M. Matuschek, *Linked Lexical Knowledge Bases: Foundations and Applications*. Morgan & Claypool Publishers, 2016. DOI: 10.2200/S00717ED1V01Y201605HLT034.

[58] J. D. Ullman and J. Widom, *A first course in database systems (2nd edition)*. Prentice Hall, 2002.

[59] W3C, *RDF Primer, https://www.w3.org/TR/2004/REC-rdf-primer-20040210/*. World Wide Web Consortium, 2004.

[60] D. Hernández, A. Hogan, and M. Krötzsch, "Reifying RDF: what works well with wikidata?", in *Scalable Semantic Web Knowledge Base Systems*, 2015. [Online]. Available: http://ceur-ws.org/Vol-1457/SSWS2015%5C_paper3.pdf.

[61] V. Christophides, V. Efthymiou, T. Palpanas, G. Papadakis, and K. Stefanidis, "End-to-end entity resolution for big data: A survey", *CoRR*, vol. abs/1905.06397, 2019.

[62] R. V. Guha and D. B. Lenat, "CYC: A mid-term report", *Appl. Artif. Intell.*, vol. 5, no. 1, 1991, pp. 45–86. DOI: 10.1080/08839519108927917.

Full text available at: http://dx.doi.org/10.1561/1900000064

[63] I. Niles and A. Pease, "Towards a standard upper ontology", in *2nd International Conference on Formal Ontology in Information Systems, FOIS 2001, Ogunquit, Maine, USA, October 17-19, 2001, Proceedings*, pp. 2–9, ACM, 2001. DOI: 10.1145/505168.505170.

[64] G. Weikum and M. Theobald, "From information to knowledge: Harvesting entities and relationships from web sources", in *ACM Symposium on Principles of Database Systems (PODS)*, 2010. DOI: 10.1145/1807085.1807097.

[65] G. Weikum, J. Hoffart, and F. M. Suchanek, "Ten years of knowledge harvesting: Lessons and challenges", *IEEE Data Engineering Bulletin*, vol. 39, no. 3, 2016, pp. 41–50. [Online]. Available: http://sites.computer.org/debull/A16sept/p41.pdf.

[66] G. Weikum, J. Hoffart, and F. M. Suchanek, "Knowledge harvesting: Achievements and challenges", in *Computing and Software Science - State of the Art and Perspectives*, ser. Lecture Notes in Computer Science, vol. 10000, Springer, 2019, pp. 217–235. DOI: 10.1007/978-3-319-91908-9_13.

[67] T. Lin, P. Pantel, M. Gamon, A. Kannan, and A. Fuxman, "Active objects: Actions for entity-centric search", in *The Web Conference (WWW)*, pp. 589–598, ACM, 2012. DOI: 10.1145/2187836.2187916.

[68] K. Chakrabarti, Z. Chen, S. Shakeri, G. Cao, and S. Chaudhuri, "Tableqna: Answering list intent queries with web tables", *CoRR*, vol. abs/2001.04828, 2020. arXiv: 2001.04828. [Online]. Available: https://arxiv.org/abs/2001.04828.

[69] L. Chiticariu, M. Danilevsky, Y. Li, F. Reiss, and H. Zhu, "Systemt: Declarative text understanding for enterprise", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2018. DOI: 10.18653/v1/n18-3010.

[70] P. M. Domingos and D. Lowd, "Unifying logical and statistical AI with markov logic", *Communication of the ACM*, vol. 62, no. 7, 2019, pp. 74–83. DOI: 10.1145/3241978.

[71] J. Eisenstein, *Introduction to Natural Language Processing*. MIT Press, 2019. [Online]. Available: https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf.

Full text available at: http://dx.doi.org/10.1561/1900000064

[72] D. Jurafsky and J. H. Martin, *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition, 3rd Edition*. Prentice Hall, Pearson Education International, 2019. [Online]. Available: https://web.stanford.edu/~jurafsky/slp3/.

[73] A. Burkov, *The Hundred-Page Machine Learning Book*. 2019. [Online]. Available: http://themlbook.com/.

[74] Y. Goldberg, *Neural Network Methods for Natural Language Processing*. Morgan & Claypool Publishers, 2017. DOI: 10.2200/S00762ED1V01Y201703HLT037.

[75] P. Ernst, A. Siu, and G. Weikum, "Knowlife: A versatile approach for constructing a large knowledge graph for biomedical sciences", *BMC Bioinformatics*, vol. 16, 2015, 157:1–157:13. DOI: 10.1186/s12859-015-0549-5.

[76] P. Ernst, A. Siu, D. Milchevski, J. Hoffart, and G. Weikum, "Deeplife: An entity-aware search, analytics and exploration platform for health and life sciences", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2016. DOI: 10.18653/v1/P16-4004.

[77] X. Ren, J. Shen, M. Qu, X. Wang, Z. Wu, Q. Zhu, M. Jiang, F. Tao, S. Sinha, D. Liem, P. Ping, R. M. Weinshilboum, and J. Han, "Life-inet: A structured network-based knowledge exploration and analytics system for life sciences", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2017. DOI: 10.18653/v1/P17-4010.

[78] D. S. Himmelstein, A. Lizee, C. Hessler, L. Brueggeman, S. L. Chen, D. Hadley, A. Green, P. Khankhanian, and S. E. Baranzini, "Systematic integration of biomedical knowledge prioritizes drugs for repurposing", *Elife*, vol. 6, 2017.

[79] Jimmy, G. Zuccon, and B. Koopman, "Choices in knowledge-base retrieval for consumer health search", in *European Conference on Information Retrieval (ECIR)*, 2018. DOI: 10.1007/978-3-319-76941-7_6.

Full text available at: http://dx.doi.org/10.1561/1900000064

[80] S. P. Ponzetto and M. Strube, "Taxonomy induction based on a collaboratively built knowledge repository", *Artificial Intelligence (AI)*, vol. 175, no. 9-10, 2011, pp. 1737–1756. DOI: 10.1016/j.artint.2011.01.003.

[81] J. Hoffart, F. M. Suchanek, K. Berberich, and G. Weikum, "YAGO2: A spatially and temporally enhanced knowledge base from wikipedia", *Artificial Intelligence (AI)*, vol. 194, 2013, pp. 28–61. DOI: 10.1016/j.artint.2012.06.001.

[82] A. Gupta, F. Piccinno, M. Kozhevnikov, M. Pasca, and D. Pighin, "Revisiting taxonomy induction over wikipedia", in *Conference on Computational Linguistics (COLING)*, 2016. [Online]. Available: http://aclweb.org/anthology/C/C16/C16-1217.pdf.

[83] M. Pasca, "Finding needles in an encyclopedic haystack: Detecting classes among wikipedia articles", in *The Web Conference (WWW)*, 2018. DOI: 10.1145/3178876.3186025.

[84] R. Navigli, "Word sense disambiguation: A survey", *ACM Computing Surveys*, vol. 41, no. 2, 2009, 10:1–10:69. DOI: 10.1145/1459352.1459355.

[85] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality", in *Neural Information Processing Systems (NeurIPS)*, 2013. [Online]. Available: http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.

[86] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014. [Online]. Available: https://www.aclweb.org/anthology/D14-1162/.

[87] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2019. [Online]. Available: https://www.aclweb.org/anthology/N19-1423/.

Full text available at: http://dx.doi.org/10.1561/1900000064

[88] J. Weeds, D. J. Weir, and D. McCarthy, "Characterising measures of lexical distributional similarity", in *Conference on Computational Linguistics (COLING)*, 2004. [Online]. Available: https://www.aclweb.org/anthology/C04-1146/.

[89] M. Geffet and I. Dagan, "The distributional inclusion hypotheses and lexical entailment", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2005. DOI: 10.3115/1219840.1219854.

[90] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine", *Computer Networks*, vol. 30, no. 1-7, 1998, pp. 107–117. DOI: 10.1016/S0169-7552(98)00110-X.

[91] G. Jeh and J. Widom, "Scaling personalized web search", in *The Web Conference (WWW)*, 2003. DOI: 10.1145/775152.775191.

[92] V. Nastase and M. Strube, "Transforming wikipedia into a large scale multilingual concept network", *Artificial Intelligence (AI)*, vol. 194, 2013, pp. 62–85. DOI: 10.1016/j.artint.2012.06.008.

[93] F. Wu and D. S. Weld, "Autonomously semantifying wikipedia", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2007. DOI: 10.1145/1321440.1321449.

[94] F. Wu and D. S. Weld, "Automatically refining the wikipedia infobox ontology", in *The Web Conference (WWW)*, 2008. DOI: 10.1145/1367497.1367583.

[95] D. Aumueller, H. H. Do, S. Massmann, and E. Rahm, "Schema and ontology matching with COMA++", in *ACM Conference on Management of Data (SIGMOD)*, 2005. DOI: 10.1145/1066157.1066283.

[96] J. Li, J. Tang, Y. Li, and Q. Luo, "Rimom: A dynamic multistrategy ontology alignment framework", *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 21, no. 8, 2009, pp. 1218–1232. DOI: 10.1109/TKDE.2008.202.

[97] W3C, *OWL 2 Web Ontology Language Document Overview (Second Edition)*, *https://www.w3.org/TR/owl2-overview/*. World Wide Web Consortium, 2012.

Full text available at: http://dx.doi.org/10.1561/1900000064

[98] O. Deshpande, D. S. Lamba, M. Tourn, S. Das, S. Subramaniam, A. Rajaraman, V. Harinarayan, and A. Doan, "Building, maintaining, and using knowledge bases: A report from the trenches", in *ACM Conference on Management of Data (SIGMOD)*, 2013.

[99] A. G. Arens-Volland, B. Gâteau, and Y. Naudet, "Semantic modeling for personalized dietary recommendation", in *IEEE Workshop on Semantic and Social Media Adaptation and Personalization*, 2018. DOI: 10.1109/SMAP.2018.8501864.

[100] S. Haussmann, O. Seneviratne, Y. Chen, Y. Ne'eman, J. Codella, C. Chen, D. L. McGuinness, and M. J. Zaki, "Foodkg: A semantics-driven knowledge graph for food recommendation", in *International Semantic Web Conference (ISWC)*, 2019. DOI: 10.1007/978-3-030-30796-7_10.

[101] K. Kari, "Building, and communicating, a knowledge graph in zalando", in *Keynote at DINAcon (Konferenz für Digitale Nachhaltigkeit), https://www.youtube.com/watch?v=QkgAFKL26Vg and https://dinacon.ch/en/interview-with-katariina-kari/*, 2019.

[102] K. Grauman, "Computer vision for fashion: From individual recommendations to world-wide trends (keynote), https://www.cs.utexas.edu/~grauman/slides/wsdm2020_fashion_grauman.pptx", in *ACM Conference on Web Search and Data Mining (WSDM)*, 2020. DOI: 10.1145/3336191.3372192.

[103] D. M. Dooley, E. J. Griffiths, G. S. Gosal, P. L. Buttigieg, R. Hoehndorf, M. C. Lange, L. M. Schriml, F. S. Brinkman, and W. W. Hsiao, "Foodon: A harmonized food ontology to increase global food traceability, quality control and data integration", *Science of Food*, vol. 2, no. 1, 2018, pp. 1–10.

[104] J. Marin, A. Biswas, F. Ofli, N. Hynes, A. Salvador, Y. Aytar, I. Weber, and A. Torralba, "Recipe1m: A dataset for learning cross-modal embeddings for cooking recipes and food images", *CoRR*, vol. abs/1810.06553, 2018.

[105] S. Hertling and H. Paulheim, "Dbkwik: A consolidated knowledge graph from thousands of wikis", in *International Conference on Big Knowledge (ICBK)*, 2018. DOI: 10.1109/ICBK.2018.00011.

Full text available at: http://dx.doi.org/10.1561/1900000064

[106] C. X. Chu, S. Razniewski, and G. Weikum, "Tifi: Taxonomy induction for fictional domains", in *The Web Conference (WWW)*, 2019. DOI: 10.1145/3308558.3313519.

[107] Z. Yang and E. Nyberg, "Leveraging procedural knowledge for task-oriented search", in *ACM Conference on Research and Development in Information Retrieval (SIGIR)*, 2015. DOI: 10.1145/2766462.2767744.

[108] C. X. Chu, N. Tandon, and G. Weikum, "Distilling task knowledge from how-to communities", in *The Web Conference (WWW)*, 2017. DOI: 10.1145/3038912.3052715.

[109] H. Cunningham, "Gate, a general architecture for text engineering", *Computers and the Humanities*, vol. 36, no. 2, 2002, pp. 223–254. DOI: 10.1023/A:1014348124664.

[110] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan, "A framework and graphical development environment for robust NLP tools and applications", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2002. DOI: 10.3115/1073083.1073112.

[111] D. A. Ferrucci and A. Lally, "UIMA: an architectural approach to unstructured information processing in the corporate research environment", *Natural Language Engineering*, vol. 10, no. 3-4, 2004, pp. 327–348. DOI: 10.1017/S1351324904003523.

[112] K. Chakrabarti, S. Chaudhuri, Z. Chen, K. Ganjam, and Y. He, "Data services leveraging bing's data assets", *IEEE Data Engineering Bulletin*, vol. 39, no. 3, 2016, pp. 15–28. [Online]. Available: http://sites.computer.org/debull/A16sept/p15.pdf.

[113] R. C. Bunescu and M. Pasca, "Using encyclopedic knowledge for named entity disambiguation", in *European Chapter of the Association for Computational Linguistics (EACL)*, 2006. [Online]. Available: https://www.aclweb.org/anthology/E06-1002/.

[114] V. I. Spitkovsky and A. X. Chang, "A cross-lingual dictionary for english wikipedia concepts", in *Conference on Language Resources and Evaluation (LREC)*, 2012. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2012/summaries/266.html.

Full text available at: http://dx.doi.org/10.1561/1900000064

[115] B. Taneva, T. Cheng, K. Chakrabarti, and Y. He, "Mining acronym expansions and their meanings using query click log", in *The Web Conference (WWW)*, 2013. DOI: 10.1145/2488388.2488498.

[116] M. Qu, X. Ren, and J. Han, "Automatic synonym discovery with knowledge bases", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2017. DOI: 10.1145/3097983.3098185.

[117] M. A. Hearst, "Automatic acquisition of hyponyms from large text corpora", in *Conference on Computational Linguistics (COLING)*, 1992. [Online]. Available: https://www.aclweb.org/anthology/C92-2082/.

[118] G. Limaye, S. Sarawagi, and S. Chakrabarti, "Annotating and searching web tables using entities, types and relationships", *Proceedings of the VLDB Endowment*, vol. 3, no. 1, 2010, pp. 1338–1347. DOI: 10.14778/1920841.1921005.

[119] Z. Kozareva and E. H. Hovy, "Learning arguments and supertypes of semantic relations using recursive patterns", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2010. [Online]. Available: https://www.aclweb.org/anthology/P10-1150/.

[120] B. B. Dalvi, W. W. Cohen, and J. Callan, "Websets: Extracting sets of entities from the web using unsupervised information extraction", in *ACM Conference on Web Search and Data Mining (WSDM)*, 2012. DOI: 10.1145/2124295.2124327.

[121] S. Brin, "Extracting patterns and relations from the world wide web", in *International Workshop on the Web and Databases (WebDB)*, 1998. DOI: 10.1007/10704656_11.

[122] D. Ravichandran and E. H. Hovy, "Learning surface text patterns for a question answering system", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2002. [Online]. Available: https://www.aclweb.org/anthology/P02-1006/.

[123] E. Agichtein and L. Gravano, "*Snowball*: Extracting relations from large plain-text collections", in *Conference on Digital Libraries*, 2000. DOI: 10.1145/336597.336644.

Full text available at: http://dx.doi.org/10.1561/1900000064

[124] R. C. Wang and W. W. Cohen, "Iterative set expansion of named entities using the web", in *IEEE Conference on Data Mining (ICDM)*, pp. 1091–1096, 2008. DOI: 10.1109/ICDM.2008.145.

[125] T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohamed, N. Nakashole, E. Platanios, A. Ritter, M. Samadi, B. Settles, R. Wang, D. Wijaya, A. Gupta, X. Chen, A. Saparov, M. Greaves, and J. Welling, "Never-ending learning", in *Conference on Artificial Intelligence (AAAI)*, 2015.

[126] R. C. Bunescu and R. J. Mooney, "A shortest path dependency kernel for relation extraction", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2005. [Online]. Available: https://www.aclweb.org/anthology/H05-1091/.

[127] A. Moschitti, D. Pighin, and R. Basili, "Semantic role labeling via tree kernel joint inference", in *Conference on Computational Natural Language Learning (CoNLL)*, 2006. [Online]. Available: https://www.aclweb.org/anthology/W06-2909/.

[128] F. M. Suchanek, G. Ifrim, and G. Weikum, "LEILA: learning to extract information by linguistic analysis", in *Ontology Learning and Population*, 2006. [Online]. Available: https://www.aclweb.org/anthology/W06-0503/.

[129] Y. He and D. Xin, "SEISA: set expansion by iterative similarity aggregation", in *The Web Conference (WWW)*, pp. 427–436, 2011. DOI: 10.1145/1963405.1963467.

[130] C. Wang, K. Chakrabarti, Y. He, K. Ganjam, Z. Chen, and P. A. Bernstein, "Concept expansion using web tables", in *The Web Conference (WWW)*, pp. 1198–1208, 2015. DOI: 10.1145/2736277.2741644.

[131] Z. Chen, M. J. Cafarella, and H. V. Jagadish, "Long-tail vocabulary dictionary extraction from the web", in *ACM Conference on Web Search and Data Mining (WSDM)*, pp. 625–634, 2016. DOI: 10.1145/2835776.2835778.

[132] A. Maedche and S. Staab, "Discovering conceptual relations from text", in *European Conference on Artificial Intelligence (ECAI)*, 2000.

Full text available at: http://dx.doi.org/10.1561/1900000064

[133] P. Pantel, D. Ravichandran, and E. H. Hovy, "Towards terascale semantic acquisition", in *Conference on Computational Linguistics (COLING)*, 2004. [Online]. Available: https://www.aclweb.org/anthology/C04-1111/.

[134] B. Aleman-Meza, C. Halaschek, A. P. Sheth, I. B. Arpinar, and G. Sannapareddy, "Sweto: Large-scale semantic web test-bed", in *Workshop on Ontology in Action, co-located with International Conference on Software Engineering and Knowledge Engineering*, 2004.

[135] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data", in *International Conference on Machine Learning (ICML)*, 2001.

[136] C. Sutton, A. McCallum, *et al.*, "An introduction to conditional random fields", *Foundations and Trends in Machine Learning*, vol. 4, no. 4, 2012, pp. 267–373.

[137] S. Sarawagi *et al.*, "Information extraction", *Foundations and Trends in Databases*, vol. 1, no. 3, 2008, pp. 261–377.

[138] D. Roth and W. Yih, "Integer linear programming inference for conditional random fields", in *International Conference on Machine Learning (ICML)*, 2005. DOI: 10.1145/1102351.1102444.

[139] J. R. Finkel, T. Grenager, and C. D. Manning, "Incorporating non-local information into information extraction systems by gibbs sampling", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2005. [Online]. Available: https://www.aclweb.org/anthology/P05-1045/.

[140] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky, "The stanford corenlp natural language processing toolkit", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2014. [Online]. Available: https://www.aclweb.org/anthology/P14-5010/.

[141] M. Fleischman and E. H. Hovy, "Fine grained classification of named entities", in *Conference on Computational Linguistics (COLING)*, 2002. [Online]. Available: https://www.aclweb.org/anthology/C02-1130/.

Full text available at: http://dx.doi.org/10.1561/1900000064

[142] X. Ling and D. S. Weld, "Fine-grained entity recognition", in *Conference on Artificial Intelligence (AAAI)*, 2012. [Online]. Available: http://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/view/5152.

[143] N. Nakashole, T. Tylenda, and G. Weikum, "Fine-grained semantic typing of emerging entities", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2013. [Online]. Available: https://www.aclweb.org/anthology/P13-1146/.

[144] E. Choi, O. Levy, Y. Choi, and L. Zettlemoyer, "Ultra-fine entity typing", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2018. DOI: 10.18653/v1/P18-1009.

[145] K. Mai, T. Pham, M. T. Nguyen, N. T. Duc, D. Bollegala, R. Sasano, and S. Sekine, "An empirical study on fine-grained named entity recognition", in *Conference on Computational Linguistics (COLING)*, 2018. [Online]. Available: https://www.aclweb.org/anthology/C18-1060/.

[146] C. S. Funk, W. A. B. Jr., B. Garcia, C. Roeder, M. Bada, K. B. Cohen, L. E. Hunter, and K. Verspoor, "Large-scale biomedical concept recognition: An evaluation of current automatic annotators and their parameters", *BMC Bioinformatics*, vol. 15, 2014, p. 59. DOI: 10.1186/1471-2105-15-59.

[147] L. Ratinov and D. Roth, "Design challenges and misconceptions in named entity recognition", in *Conference on Computational Natural Language Learning (CoNLL)*, 2009. [Online]. Available: https://www.aclweb.org/anthology/W09-1119/.

[148] J. Shang, L. Liu, X. Gu, X. Ren, T. Ren, and J. Han, "Learning named entity tagger using domain-specific dictionary", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018. [Online]. Available: https://www.aclweb.org/anthology/D18-1230/.

[149] J. Schmidhuber, "Deep learning in neural networks: An overview", *Neural Networks*, vol. 61, 2015, pp. 85–117. DOI: 10.1016/j.neunet.2014.09.003.

[150] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging", *CoRR abs/1508.01991*, 2015.

Full text available at: http://dx.doi.org/10.1561/1900000064

[151] X. Ma and E. H. Hovy, "End-to-end sequence labeling via bi-directional lstm-cnns-crf", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2016. [Online]. Available: https://www.aclweb.org/anthology/P16-1101/.

[152] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural architectures for named entity recognition", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2016. [Online]. Available: https://www.aclweb.org/anthology/N16-1030/.

[153] J. Li, A. Sun, J. Han, and C. Li, "A survey on deep learning for named entity recognition", *CoRR*, vol. abs/1812.09449, 2018.

[154] O. Levy, Y. Goldberg, and I. Dagan, "Improving distributional similarity with lessons learned from word embeddings", *Transactions of the Association for Computational Linguistics (TACL)*, vol. 3, 2015, pp. 211–225. [Online]. Available: https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/570.

[155] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman, "Indexing by latent semantic analysis", *JASIS*, vol. 41, no. 6, 1990, pp. 391–407.

[156] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis", *Machine Learning*, vol. 42, no. 1/2, 2001, pp. 177–196. DOI: 10.1023/A:1007617005950.

[157] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation", *Journal of Machine Learning Research (JMLR)*, vol. 3, 2003, pp. 993–1022. [Online]. Available: http://jmlr.org/papers/v3/blei03a.html.

[158] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2018. [Online]. Available: https://www.aclweb.org/anthology/N18-1202/.

[159] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized BERT pretraining approach", *CoRR*, vol. abs/1907.11692, 2019. arXiv: 1907.11692. [Online]. Available: http://arxiv.org/abs/1907.11692.

Full text available at: http://dx.doi.org/10.1561/1900000064

[160] A. Rogers, O. Kovaleva, and A. Rumshisky, "A primer in bertology: What we know about how BERT works", *Transactions of the Association for Computational Linguistics (TACL)*, vol. 8, 2020, pp. 842–866. [Online]. Available: https://transacl.org/ojs/index.php/tacl/article/view/2257.

[161] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph and text jointly embedding", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014. [Online]. Available: https://www.aclweb.org/anthology/D14-1167/.

[162] S. Zwicklbauer, C. Seifert, and M. Granitzer, "Robust and collective entity disambiguation through semantic embeddings", in *ACM Conference on Research and Development in Information Retrieval (SIGIR)*, 2016. DOI: 10.1145/2911451.2911535.

[163] I. Yamada, A. Asai, H. Shindo, H. Takeda, and Y. Takefuji, "Wikipedia2vec: An optimized tool for learning embeddings of words and entities from wikipedia", *CoRR*, vol. abs/1812.06280, 2018.

[164] Y. Zhang, Q. Chen, Z. Yang, H. Lin, and Z. Lu, "Biowordvec, improving biomedical word embeddings with subword information and mesh", *Scientific Data*, vol. 6, 2019, p. 52. DOI: 10.1038/s41597-019-0055-0.

[165] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, "Biobert: A pre-trained biomedical language representation model for biomedical text mining", *Bioinformatics*, vol. 36, no. 4, Sep. 2019, pp. 1234–1240. DOI: 10.1093/bioinformatics/btz682.

[166] E. Gabrilovich and S. Markovitch, "Computing semantic relatedness using wikipedia-based explicit semantic analysis", in *Joint Conference on Artificial Intelligence (IJCAI)*, 2007. [Online]. Available: http://ijcai.org/Proceedings/07/Papers/259.pdf.

[167] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge graph embedding: A survey of approaches and applications", *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 29, no. 12, 2017, pp. 2724–2743. DOI: 10.1109/TKDE.2017.2754499.

Full text available at: http://dx.doi.org/10.1561/1900000064

[168] R. Snow, D. Jurafsky, and A. Y. Ng, "Semantic taxonomy induction from heterogenous evidence", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2006. [Online]. Available: https://www.aclweb.org/anthology/P06-1101/.

[169] C. Wang, M. Danilevsky, J. Liu, N. Desai, H. Ji, and J. Han, "Constructing topical hierarchies in heterogeneous information networks", in *IEEE Conference on Data Mining (ICDM)*, pp. 767–776, 2013. DOI: 10.1109/ICDM.2013.53.

[170] J. Shang, X. Zhang, L. Liu, S. Li, and J. Han, "Nettaxo: Automated topic taxonomy construction from text-rich network", in *The Web Conference (WWW)*, pp. 1908–1919, 2020. DOI: 10.1145/3366423.3380259.

[171] M. Gupta, R. Li, Z. Yin, and J. Han, "Survey on social tagging techniques", *SIGKDD Explorations*, vol. 12, no. 1, 2010, pp. 58–72. DOI: 10.1145/1882471.1882480.

[172] P. Heymann and H. Garcia-Molina, "Collaborative creation of communal hierarchical taxonomies in social tagging systems", *Stanford University, Technical Report*, 2006.

[173] A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme, "Information retrieval in folksonomies: Search and ranking", in *European Semantic Web Conference (ESWC)*, 2006. DOI: 10.1007/11762256_31.

[174] R. Jäschke, A. Hotho, C. Schmitz, B. Ganter, and G. Stumme, "Discovering shared conceptualizations in folksonomies", *Journal of Web Semantics (JWS)*, vol. 6, no. 1, 2008, pp. 38–53. DOI: 10.1016/j.websem.2007.11.004.

[175] B. V. Durme and M. Pasca, "Finding cars, goddesses and enzymes: Parametrizable acquisition of labeled instances for open-domain information extraction", in *Conference on Artificial Intelligence (AAAI)*, 2008. [Online]. Available: http://www.aaai.org/Library/AAAI/2008/aaai08-197.php.

Full text available at: http://dx.doi.org/10.1561/1900000064

[176] P. P. Talukdar, J. Reisinger, M. Pasca, D. Ravichandran, R. Bhagat, and F. C. N. Pereira, "Weakly-supervised acquisition of labeled class instances using graph random walks", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2008. [Online]. Available: https://www.aclweb.org/anthology/D08-1061/.

[177] M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. M. Mitchell, K. Nigam, and S. Slattery, "Learning to construct knowledge bases from the world wide web", *Artificial Intelligence (AI)*, vol. 118, no. 1-2, 2000, pp. 69–113. DOI: 10.1016/S0004-3702(00)00004-7.

[178] M. Bansal, D. Burkett, G. de Melo, and D. Klein, "Structured learning for taxonomy induction with belief propagation", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 1041–1051, 2014. DOI: 10.3115/v1/p14-1098.

[179] V. Shwartz, Y. Goldberg, and I. Dagan, "Improving hypernymy detection with an integrated path-based and distributional method", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2016. DOI: 10.18653/v1/p16-1226.

[180] Y. Mao, X. Ren, J. Shen, X. Gu, and J. Han, "End-to-end reinforcement learning for automatic taxonomy induction", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 2462–2472, 2018. DOI: 10.18653/v1/P18-1229.

[181] R. A. Baeza-Yates and A. Tiberi, "Extracting semantic relations from query logs", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2007. DOI: 10.1145/1281192.1281204.

[182] M. Pasca, "Open-domain fine-grained class extraction from web search queries", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2013. [Online]. Available: https://www.aclweb.org/anthology/D13-1039/.

[183] B. Liu, W. Guo, D. Niu, C. Wang, S. Xu, J. Lin, K. Lai, and Y. Xu, "A user-centered concept mining system for query and document understanding at tencent", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2019. DOI: 10.1145/3292500.3330727.

Full text available at: http://dx.doi.org/10.1561/1900000064

[184] B. Liu, W. Guo, D. Niu, J. Luo, C. Wang, Z. Wen, and Y. Xu, "GIANT: scalable creation of a web-scale ontology", in *ACM Conference on Management of Data (SIGMOD)*, ACM, 2020. DOI: 10.1145/3318464.3386145.

[185] B. Koopman and G. Zuccon, "WSDM 2019 tutorial on health search (HS2019): A full-day from consumers to clinicians (with materials on https://github.com/ielab/health-search-tutorial/tree/wsdm2019)", in *ACM Conference on Web Search and Data Mining (WSDM)*, pp. 838–839, 2019. DOI: 10.1145/3289600.3291379.

[186] R. Mihalcea and A. Csomai, "Wikify!: Linking documents to encyclopedic knowledge", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2007. DOI: 10.1145/1321440.1321475.

[187] D. N. Milne and I. H. Witten, "Learning to link with wikipedia", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2008. DOI: 10.1145/1458082.1458150.

[188] A. Moro, A. Raganato, and R. Navigli, "Entity linking meets word sense disambiguation: A unified approach", *Transactions of the Association for Computational Linguistics (TACL)*, vol. 2, 2014, pp. 231–244. [Online]. Available: https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/291.

[189] X. Ling, S. Singh, and D. S. Weld, "Design challenges for entity linking", *Transactions of the Association for Computational Linguistics (TACL)*, vol. 3, 2015, pp. 315–328. [Online]. Available: https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/528.

[190] J. Martinez-Rodriguez, A. Hogan, and I. Lopez-Arvalo, "Information extraction meets the semantic web: A survey", *Semantic Web Journal (SWJ)*, vol. 11, no. 2, 2020, pp. 255–335. DOI: 10.3233/SW-180333.

[191] B. Hachey, W. Radford, J. Nothman, M. Honnibal, and J. R. Curran, "Evaluating entity linking with wikipedia", *Artificial Intelligence (AI)*, vol. 194, 2013, pp. 130–150. DOI: 10.1016/j.artint.2012.04.005.

Full text available at: http://dx.doi.org/10.1561/1900000064

[192] M. Röder, R. Usbeck, and A. N. Ngomo, "GERBIL - benchmarking named entity recognition and linking consistently", *Semantic Web Journal (SWJ)*, vol. 9, no. 5, 2018, pp. 605–625. DOI: 10.3233/SW-170286.

[193] K. Raghunathan, H. Lee, S. Rangarajan, N. Chambers, M. Surdeanu, D. Jurafsky, and C. D. Manning, "A multi-pass sieve for coreference resolution", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2010. [Online]. Available: https://www.aclweb.org/anthology/D10-1048/.

[194] H. Lee, A. X. Chang, Y. Peirsman, N. Chambers, M. Surdeanu, and D. Jurafsky, "Deterministic coreference resolution based on entity-centric, precision-ranked rules", *Computational Linguistics*, vol. 39, no. 4, 2013, pp. 885–916. DOI: 10.1162/COLI_a_00152.

[195] G. Durrett and D. Klein, "Easy victories and uphill battles in coreference resolution", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2013. [Online]. Available: https://www.aclweb.org/anthology/D13-1203/.

[196] G. Durrett and D. Klein, "A joint model for entity analysis: Coreference, typing, and linking", *Transactions of the Association for Computational Linguistics (TACL)*, vol. 2, 2014, pp. 477–490. [Online]. Available: https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/412.

[197] K. Clark and C. D. Manning, "Improving coreference resolution by learning entity-level distributed representations", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2016. DOI: 10.18653/v1/p16-1061.

[198] K. Lee, L. He, M. Lewis, and L. Zettlemoyer, "End-to-end neural coreference resolution", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2017. DOI: 10.18653/v1/d17-1018.

[199] M. Joshi, O. Levy, L. Zettlemoyer, and D. S. Weld, "BERT for coreference resolution: Baselines and analysis", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019. DOI: 10.18653/v1/D19-1588.

Full text available at: http://dx.doi.org/10.1561/1900000064

[200] A. Bagga and B. Baldwin, "Entity-based cross-document coreferencing using the vector space model", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 1998. [Online]. Available: https://www.aclweb.org/anthology/P98-1012/.

[201] S. Singh, A. Subramanya, F. C. N. Pereira, and A. McCallum, "Large-scale cross-document coreference using distributed inference and hierarchical models", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2011. [Online]. Available: https://www.aclweb.org/anthology/P11-1080/.

[202] S. Dutta and G. Weikum, "C3EL: A joint model for cross-document co-reference resolution and entity linking", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2015. DOI: 10.18653/v1/d15-1101.

[203] H. L. Dunn, "Record linkage", *American Journal of Public Health and the Nations Health*, vol. 36, no. 12, 1946, pp. 1412–1416.

[204] I. P. Fellegi and A. B. Sunter, "A theory for record linkage", *Journal of the American Statistical Association*, vol. 64, no. 328, 1969, pp. 1183–1210.

[205] P. Konda, S. Das, P. S. G. C., A. Doan, A. Ardalan, J. R. Ballard, H. Li, F. Panahi, H. Zhang, J. F. Naughton, S. Prasad, G. Krishnan, R. Deep, and V. Raghavendra, "Magellan: Toward building entity matching management systems", *Proceedings of the VLDB Endowment*, vol. 9, no. 12, 2016, pp. 1197–1208. DOI: 10.14778/2994509.2994535.

[206] R. Singh, V. V. Meduri, A. K. Elmagarmid, S. Madden, P. Papotti, J. Quiané-Ruiz, A. Solar-Lezama, and N. Tang, "Synthesizing entity matching rules by examples", *Proceedings of the VLDB Endowment*, vol. 11, no. 2, 2017, pp. 189–202. DOI: 10.14778/3149193.3149199.

[207] P. Singla and P. M. Domingos, "Entity resolution with markov logic", in *IEEE Conference on Data Mining (ICDM)*, 2006. DOI: 10.1109/ICDM.2006.65.

[208] I. Bhattacharya and L. Getoor, "Collective entity resolution in relational data", *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 1, no. 1, 2007, p. 5. DOI: 10.1145/1217299.1217304.

Full text available at: http://dx.doi.org/10.1561/1900000064

[209] M. L. Wick, K. Rohanimanesh, K. Schultz, and A. McCallum, "A unified approach for schema matching, coreference and canonicalization", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2008. DOI: 10.1145/1401890.1401977.

[210] V. Rastogi, N. N. Dalvi, and M. N. Garofalakis, "Large-scale collective entity matching", *Proceedings of the VLDB Endowment*, vol. 4, no. 4, 2011, pp. 208–218. DOI: 10.14778/1938545.1938546.

[211] F. M. Suchanek, S. Abiteboul, and P. Senellart, "PARIS: probabilistic alignment of relations, instances, and schema", *Proceedings of the VLDB Endowment*, vol. 5, no. 3, 2011, pp. 157–168. DOI: 10.14778/2078331.2078332.

[212] P. Kouki, J. Pujara, C. Marcum, L. M. Koehly, and L. Getoor, "Collective entity resolution in multi-relational familial networks", *Knowledge and Information Systems (KAIS)*, vol. 61, no. 3, 2019, pp. 1547–1581. DOI: 10.1007/s10115-018-1246-2.

[213] S. Mudgal, H. Li, T. Rekatsinas, A. Doan, Y. Park, G. Krishnan, R. Deep, E. Arcaute, and V. Raghavendra, "Deep learning for entity matching: A design space exploration", in *ACM Conference on Management of Data (SIGMOD)*, 2018. DOI: 10.1145/3183713.3196926.

[214] Q. Zhu, H. Wei, B. Sisman, D. Zheng, C. Faloutsos, X. L. Dong, and J. Han, "Collective multi-type entity alignment between knowledge graphs", in *The Web Conference (WWW)*, pp. 2241–2252, 2020. DOI: 10.1145/3366423.3380289.

[215] R. J. Miller, F. Nargesian, E. Zhu, C. Christodoulakis, K. Q. Pu, and P. Andritsos, "Making open data transparent: Data discovery on open data", *IEEE Data Engineering Bulletin*, vol. 41, no. 2, 2018, pp. 59–70. [Online]. Available: http://sites.computer.org/debull/A18june/p59.pdf.

[216] M. Nentwig, M. Hartung, A. N. Ngomo, and E. Rahm, "A survey of current link discovery frameworks", *Semantic Web Journal (SWJ)*, vol. 8, no. 3, 2017, pp. 419–436. DOI: 10.3233/SW-150210.

[217] E. Zhu, D. Deng, F. Nargesian, and R. J. Miller, "JOSIE: overlap set similarity search for finding joinable tables in data lakes", in *ACM Conference on Management of Data (SIGMOD)*, 2019. DOI: 10.1145/3299869.3300065.

Full text available at: http://dx.doi.org/10.1561/1900000064

[218] O. Lehmberg and C. Bizer, "Stitching web tables for improving matching quality", *Proceedings of the VLDB Endowment*, vol. 10, no. 11, 2017, pp. 1502–1513. DOI: 10.14778/3137628.3137657.

[219] H. Köpcke and E. Rahm, "Frameworks for entity matching: A comparison", *Data & Knowledge Engineering*, vol. 69, no. 2, 2010, pp. 197–210. DOI: 10.1016/j.datak.2009.10.003.

[220] F. Naumann and M. Herschel, *An Introduction to Duplicate Detection*. Morgan & Claypool Publishers, 2010. DOI: 10.2200/S00262ED1V01Y201003DTM003.

[221] P. Christen, *Data Matching - Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Springer, 2012. DOI: 10.1007/978-3-642-31164-2.

[222] X. L. Dong and D. Srivastava, *Big Data Integration*. Morgan & Claypool Publishers, 2015. DOI: 10.2200/S00578ED1V01Y201404DTM040.

[223] O. Medelyan, I. H. Witten, and D. Milne, "Topic indexing with wikipedia", in *AAAI WikiAI workshop*, pp. 19–24, 2008.

[224] S. Dill, N. Eiron, D. Gibson, D. Gruhl, R. V. Guha, A. Jhingran, T. Kanungo, S. Rajagopalan, A. Tomkins, J. A. Tomlin, and J. Y. Zien, "Semtag and seeker: Bootstrapping the semantic web via automated semantic annotation", in *The Web Conference (WWW)*, 2003. DOI: 10.1145/775152.775178.

[225] S. Cucerzan, "Large-scale named entity disambiguation based on wikipedia data", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2007. [Online]. Available: https://www.aclweb.org/anthology/D07-1074/.

[226] P. N. Mendes, M. Jakob, A. Garcia-Silva, and C. Bizer, "Dbpedia spotlight: Shedding light on the web of documents", in *International Conference on Semantic Systems*, 2011. DOI: 10.1145/2063518.2063519.

[227] J. Hoffart, S. Seufert, D. B. Nguyen, M. Theobald, and G. Weikum, "KORE: keyphrase overlap relatedness for entity disambiguation", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2012. DOI: 10.1145/2396761.2396832.

Full text available at: http://dx.doi.org/10.1561/1900000064

[228] S. Kulkarni, A. Singh, G. Ramakrishnan, and S. Chakrabarti, "Collective annotation of wikipedia entities in web text", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2009. DOI: 10.1145/1557019.1557073.

[229] P. Ferragina and U. Scaiella, "TAGME: on-the-fly annotation of short text fragments (by wikipedia entities)", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2010. DOI: 10.1145/1871437.1871689.

[230] X. Han, L. Sun, and J. Zhao, "Collective entity linking in web text: A graph-based method", in *ACM Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 765–774, ACM, 2011. DOI: 10.1145/2009916.2010019.

[231] L. Ratinov, D. Roth, D. Downey, and M. Anderson, "Local and global algorithms for disambiguation to wikipedia", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2011. [Online]. Available: https://www.aclweb.org/anthology/P11-1138/.

[232] J. Hoffart, M. A. Yosef, I. Bordino, H. Fürstenau, M. Pinkal, M. Spaniol, B. Taneva, S. Thater, and G. Weikum, "Robust disambiguation of named entities in text", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2011. [Online]. Available: https://www.aclweb.org/anthology/D11-1072/.

[233] Y. Li, S. Tan, H. Sun, J. Han, D. Roth, and X. Yan, "Entity disambiguation with linkless knowledge bases", in *The Web Conference (WWW)*, 2016. DOI: 10.1145/2872427.2883068.

[234] Z. Guo and D. Barbosa, "Robust named entity disambiguation with random walks", *Semantic Web Journal (SWJ)*, vol. 9, no. 4, 2018, pp. 459–479. DOI: 10.3233/SW-170273.

[235] M. Strube and S. P. Ponzetto, "Wikirelate! computing semantic relatedness using wikipedia", in *National Conference on Artificial Intelligence*, 2006. [Online]. Available: http://www.aaai.org/Library/AAAI/2006/aaai06-223.php.

Full text available at: http://dx.doi.org/10.1561/1900000064

[236] D. Ceccarelli, C. Lucchese, S. Orlando, R. Perego, and S. Trani, "Learning relatedness measures for entity linking", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2013. DOI: 10.1145/2505515.2505711.

[237] M. Ponza, P. Ferragina, and S. Chakrabarti, "On computing entity relatedness in wikipedia, with applications", *Knowledge Based Systems*, vol. 188, 2020. DOI: 10.1016/j.knosys.2019.105051.

[238] M. Dredze, P. McNamee, D. Rao, A. Gerber, and T. Finin, "Entity disambiguation for knowledge base population", in *Conference on Computational Linguistics (COLING)*, 2010. [Online]. Available: https://www.aclweb.org/anthology/C10-1032/.

[239] W. Shen, J. Wang, P. Luo, and M. Wang, "LINDEN: linking named entities with knowledge base via semantic knowledge", in *The Web Conference (WWW)*, 2012. DOI: 10.1145/2187836.2187898.

[240] N. Lazic, A. Subramanya, M. Ringgaard, and F. Pereira, "Plato: A selective context model for entity resolution", *Transactions of the Association for Computational Linguistics (TACL)*, vol. 3, 2015, pp. 503–515. [Online]. Available: https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/637.

[241] T. Liu, *Learning to Rank for Information Retrieval*. Springer, 2011. DOI: 10.1007/978-3-642-14267-3.

[242] H. Li, *Learning to Rank for Information Retrieval and Natural Language Processing, Second Edition*. Morgan & Claypool Publishers, 2014. DOI: 10.2200/S00607ED2V01Y201410HLT026.

[243] Z. Zheng, F. Li, M. Huang, and X. Zhu, "Learning to link entities with knowledge base", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2010. [Online]. Available: https://www.aclweb.org/anthology/N10-1072/.

[244] S. Cucerzan and A. Sil, "The MSR systems for entity linking and temporal slot filling at TAC 2013", in *Text Analysis Conference (TAC)*, 2013.

Full text available at: http://dx.doi.org/10.1561/1900000064

[245] T. H. Haveliwala, "Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search", *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 15, no. 4, 2003, pp. 784–796. DOI: 10.1109/TKDE.2003.1208999.

[246] Z. Guo and D. Barbosa, "Robust entity linking via random walks", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2014. DOI: 10.1145/2661829.2661887.

[247] F. Piccinno and P. Ferragina, "From tagme to WAT: a new entity annotator", in *International Workshop on Entity Recognition & Disambiguation, co-located with ACM SIGIR Conference*, 2014. DOI: 10.1145/2633211.2634350.

[248] T. Grütze, G. Kasneci, Z. Zuo, and F. Naumann, "Coheel: Coherent and efficient named entity linking through random walks", *Journal of Web Semantics (JWS)*, vol. 37-38, 2016, pp. 75–89. DOI: 10.1016/j.websem.2016.03.001.

[249] O. Ganea, M. Ganea, A. Lucchi, C. Eickhoff, and T. Hofmann, "Probabilistic bag-of-hyperlinks model for entity linking", in *The Web Conference (WWW)*, 2016. DOI: 10.1145/2872427.2882988.

[250] D. B. Nguyen, M. Theobald, and G. Weikum, "J-NERD: joint named entity recognition and disambiguation with rich linguistic features", *Transactions of the Association for Computational Linguistics (TACL)*, vol. 4, 2016, pp. 215–229. [Online]. Available: https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/698.

[251] D. Roth and V. Srikumar, "Integer linear programming formulations in natural language processing (tutorial materials)", in *Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 2017. [Online]. Available: https://ilpinference.github.io/eacl2017/.

[252] Z. He, S. Liu, M. Li, M. Zhou, L. Zhang, and H. Wang, "Learning entity representation for entity disambiguation", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 30–34, The Association for Computer Linguistics, 2013. [Online]. Available: https://www.aclweb.org/anthology/P13-2006/.

Full text available at: http://dx.doi.org/10.1561/1900000064

[253] Y. Sun, L. Lin, D. Tang, N. Yang, Z. Ji, and X. Wang, "Modeling mention, context and entity with neural networks for entity disambiguation", in *Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1333–1339, AAAI Press, 2015. [Online]. Available: http://ijcai.org/Abstract/15/192.

[254] M. Francis-Landau, G. Durrett, and D. Klein, "Capturing semantic similarity for entity linking with convolutional neural networks", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2016. DOI: 10.18653/v1/n16-1150.

[255] I. Yamada, H. Shindo, H. Takeda, and Y. Takefuji, "Joint learning of the embedding of words and entities for named entity disambiguation", in *Conference on Computational Natural Language Learning (CoNLL)*, 2016. DOI: 10.18653/v1/k16-1025.

[256] I. Yamada, H. Shindo, H. Takeda, and Y. Takefuji, "Learning distributed representations of texts and entities from knowledge base", *Transactions of the Association for Computational Linguistics (TACL)*, vol. 5, 2017, pp. 397–411. [Online]. Available: https://transacl.org/ojs/index.php/tacl/article/view/1065.

[257] N. Gupta, S. Singh, and D. Roth, "Entity linking via joint encoding of types, descriptions, and context", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2017. DOI: 10.18653/v1/d17-1284.

[258] N. Kolitsas, O. Ganea, and T. Hofmann, "End-to-end neural entity linking", in *Conference on Computational Natural Language Learning (CoNLL)*, 2018. DOI: 10.18653/v1/k18-1050.

[259] Y. Eshel, N. Cohen, K. Radinsky, S. Markovitch, I. Yamada, and O. Levy, "Named entity disambiguation for noisy text", in *Conference on Computational Natural Language Learning (CoNLL)*, 2017. DOI: 10.18653/v1/K17-1008.

[260] D. Mueller and G. Durrett, "Effective use of context in noisy entity linking", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018. DOI: 10.18653/v1/d18-1126.

Full text available at: http://dx.doi.org/10.1561/1900000064

[261] özge Sevgili, A. Panchenko, and C. Biemann, "Improving neural entity disambiguation with graph embeddings", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019. DOI: 10.18653/v1/p19-2044.

[262] J. M. van Hulst, F. Hasibi, K. Dercksen, K. Balog, and A. P. de Vries, "REL: an entity linker standing on the shoulders of giants", in *ACM Conference on Research and Development in Information Retrieval (SIGIR)*, 2020.

[263] M. P. K. Ravi, K. Singh, I. O. Mulang, S. Shekarpour, J. Hoffart, and J. Lehmann, "CHOLAN: A modular approach for neural entity linking on wikipedia and wikidata", in *European Chapter of the Association for Computational Linguistics (EACL)*, pp. 504–514, Association for Computational Linguistics, 2021. [Online]. Available: https://www.aclweb.org/anthology/2021.eacl-main.40/.

[264] Ö. Sevgili, A. Shelmanov, M. Arkhipov, A. Panchenko, and C. Biemann, "Neural entity linking: A survey of models based on deep learning", *CoRR*, vol. abs/2006.00575, 2020. arXiv: 2006.00575. [Online]. Available: https://arxiv.org/abs/2006.00575.

[265] S. Singh, A. Subramanya, F. Pereira, and A. McCallum, "Wikilinks: A large-scale cross-document coreference corpus labeled via links to wikipedia", *University of Massachusetts, Amherst, Technical Report UM-CS-2012*, vol. 15, 2012.

[266] R. Kar, S. Reddy, S. Bhattacharya, A. Dasgupta, and S. Chakrabarti, "Task-specific representation learning for web-scale entity disambiguation", in *Conference on Artificial Intelligence (AAAI)*, S. A. McIlraith and K. Q. Weinberger, Eds., pp. 5812–5819, AAAI Press, 2018. [Online]. Available: https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17281.

[267] L. Logeswaran, M. Chang, K. Lee, K. Toutanova, J. Devlin, and H. Lee, "Zero-shot entity linking by reading entity descriptions", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019. DOI: 10.18653/v1/p19-1335.

[268] L. Wu, F. Petroni, M. Josifoski, S. Riedel, and L. Zettlemoyer, "Zero-shot entity linking with dense entity retrieval", *CoRR*, vol. abs/1911.03814, 2019.

Full text available at: http://dx.doi.org/10.1561/1900000064

[269] L. Wu, F. Petroni, M. Josifoski, S. Riedel, and L. Zettlemoyer, "Scalable zero-shot entity linking with dense entity retrieval", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6397–6407, Association for Computational Linguistics, 2020. DOI: 10.18653/v1/2020.emnlp-main.519.

[270] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners", *arXiv: 2005.14165*, 2020.

[271] P. Venetis, A. Y. Halevy, J. Madhavan, M. Pasca, W. Shen, F. Wu, G. Miao, and C. Wu, "Recovering semantics of tables on the web", *Proceedings of the VLDB Endowment*, vol. 4, no. 9, 2011, pp. 528–538. DOI: 10.14778/2002938.2002939.

[272] J. Chen, E. Jiménez-Ruiz, I. Horrocks, and C. A. Sutton, "Colnet: Embedding the semantics of web tables for column type prediction", in *Conference on Artificial Intelligence (AAAI)*, 2019. DOI: 10.1609/aaai.v33i01.330129.

[273] W. Shen, J. Wang, P. Luo, and M. Wang, "LIEGE: : Link entities in web lists with knowledge base", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2012. DOI: 10.1145/2339530.2339753.

[274] H. Paulheim and S. P. Ponzetto, "Extending dbpedia with wikipedia list pages", in *NLP & DBpedia workshop, co-located with International Semantic Web Conference (ISWC 2013)*, ser. CEUR Workshop Proceedings, vol. 1064, CEUR-WS.org, 2013. [Online]. Available: http://ceur-ws.org/Vol-1064/Paulheim%5C_Extending%5C_DBpedia.pdf.

[275] N. Heist and H. Paulheim, "Entity extraction from wikipedia list pages", in *European Semantic Web Conference (ESWC)*, 2020. DOI: 10.1007/978-3-030-49461-2_19.

Full text available at: http://dx.doi.org/10.1561/1900000064

[276] C. S. Bhagavatula, T. Noraset, and D. Downey, "Tabel: Entity linking in web tables", in *International Semantic Web Conference (ISWC)*, 2015. DOI: 10.1007/978-3-319-25007-6_25.

[277] Y. Ibrahim, M. Riedewald, and G. Weikum, "Making sense of entities and quantities in web tables", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2016. DOI: 10.1145/2983323.2983772.

[278] D. Ritze and C. Bizer, "Matching web tables to dbpedia - A feature utility study", in *Conference on Extending Database Technology (EDBT)*, 2017. DOI: 10.5441/002/edbt.2017.20.

[279] S. Zhang and K. Balog, "Web table extraction, retrieval, and augmentation: A survey", *ACM Transactions on Intelligent Systems Technology (TIST)*, vol. 11, no. 2, 2020, 13:1–13:35. DOI: 10.1145/3372117.

[280] D. B. Nguyen, J. Hoffart, M. Theobald, and G. Weikum, "Aida-light: High-throughput named-entity disambiguation", in *Workshop on Linked Data on the Web (LDOW), co-located with the Web Conference*, 2014. [Online]. Available: http://ceur-ws.org/Vol-1184/ldow2014%5C_paper%5C_03.pdf.

[281] A. R. Aronson and F. Lang, "An overview of metamap: Historical perspective and recent advances", *J. Am. Medical Informatics Assoc.*, vol. 17, no. 3, 2010, pp. 229–236. DOI: 10.1136/jamia.2009.002733.

[282] J. Dai, M. Zhang, G. Chen, J. Fan, K. Y. Ngiam, and B. C. Ooi, "Fine-grained concept linking using neural networks in healthcare", in *ACM Conference on Management of Data (SIGMOD)*, ACM, 2018. DOI: 10.1145/3183713.3196907.

[283] Z. Ji, Q. Wei, and H. Xu, "Bert-based ranking for biomedical entity normalization", *CoRR*, vol. abs/1908.03548, 2019.

[284] A. Sil, H. Ji, D. Roth, and S. Cucerzan, "Multi-lingual entity discovery and linking", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2018. DOI: 10.18653/v1/P18-5008.

Full text available at: http://dx.doi.org/10.1561/1900000064

[285] E. Meij, W. Weerkamp, and M. de Rijke, "Adding semantics to microblog posts", in *ACM Conference on Web Search and Data Mining (WSDM)*, pp. 563–572, ACM, 2012. DOI: 10.1145/2124295.2124364.

[286] X. Liu, Y. Li, H. Wu, M. Zhou, F. Wei, and Y. Lu, "Entity linking for tweets", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2013. [Online]. Available: https://www.aclweb.org/anthology/P13-1128/.

[287] W. Shen, J. Wang, P. Luo, and M. Wang, "Linking named entities in tweets with knowledge base via user interest modeling", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2013. DOI: 10.1145/2487575.2487686.

[288] L. Derczynski, D. Maynard, G. Rizzo, M. van Erp, G. Gorrell, R. Troncy, J. Petrak, and K. Bontcheva, "Analysis of named entity recognition and linking for tweets", *Information Processing and Management*, vol. 51, no. 2, 2015, pp. 32–49. DOI: 10.1016/j.ipm.2014.10.006.

[289] U. Sawant and S. Chakrabarti, "Learning joint query interpretation and response ranking", in *The Web Conference (WWW)*, 2013. DOI: 10.1145/2488388.2488484.

[290] M. Yahya, K. Berberich, S. Elbassuoni, M. Ramanath, V. Tresp, and G. Weikum, "Natural language questions for the web of data", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 379–390, ACL, 2012. [Online]. Available: https://www.aclweb.org/anthology/D12-1035/.

[291] R. Blanco, G. Ottaviano, and E. Meij, "Fast and space-efficient entity linking for queries", in *ACM Conference on Web Search and Data Mining (WSDM)*, pp. 179–188, ACM, 2015. DOI: 10.1145/2684822.2685317.

[292] B. Z. Li, S. Min, S. Iyer, Y. Mehdad, and W. Yih, "Efficient one-pass end-to-end entity linking for questions", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, B. Webber, T. Cohn, Y. He, and Y. Liu, Eds., pp. 6433–6441, Association for Computational Linguistics, 2020. DOI: 10.18653/v1/2020.emnlp-main.522.

Full text available at: http://dx.doi.org/10.1561/1900000064

[293] J. L. Leidner, *Toponym resolution in text: Annotation, evaluation and applications of spatial grounding of place names*. Universal-Publishers, 2008.

[294] H. Samet, J. Sankaranarayanan, M. D. Lieberman, M. D. Adelfio, B. C. Fruin, J. M. Lotkowski, D. Panozzo, J. Sperling, and B. E. Teitler, "Reading news with maps by exploiting spatial synonyms", *Communication of the ACM*, vol. 57, no. 10, 2014, pp. 64–77. DOI: 10.1145/2629572.

[295] H. Chen, M. Vasardani, and S. Winter, "Clustering-based disambiguation of fine-grained place names from descriptions", *GeoInformatica*, vol. 23, no. 3, 2019, pp. 449–472. DOI: 10.1007/s10707-019-00341-6.

[296] J. Strötgen, M. Gertz, G. Hirst, and R. Huang, "Domain-sensitive temporal tagging", *Computational Linguistics*, vol. 44, no. 2, 2018. DOI: 10.1162/coli_r_00319.

[297] E. Kuzey, V. Setty, J. Strötgen, and G. Weikum, "As time goes by: Comprehensive tagging of textual phrases with temporal scopes", in *The Web Conference (WWW)*, 2016. DOI: 10.1145/2872427.2883055.

[298] K. Leetaru and P. A. Schrodt, "Gdelt: Global data on events, location, and tone, 1979–2012", in *ISA Annual Convention*, pp. 1–49, 2013.

[299] M. Stonebraker and I. F. Ilyas, "Data integration: The current status and the way forward", *IEEE Data Engineering Bulletin*, vol. 41, no. 2, 2018, pp. 3–9. [Online]. Available: http://sites.computer.org/debull/A18june/p3.pdf.

[300] H. B. Newcombe, "Record linking: The design of efficient systems for linking records into individual and family histories", *American Journal of Human Genetics*, vol. 19, no. 3 Pt 1, 1967, p. 335.

[301] A. Z. Broder, S. C. Glassman, M. S. Manasse, and G. Zweig, "Syntactic clustering of the web", *Computer Networks*, vol. 29, no. 8-13, 1997, pp. 1157–1166. DOI: 10.1016/S0169-7552(97)00031-7.

Full text available at: http://dx.doi.org/10.1561/1900000064

[302] A. Z. Broder, M. Charikar, A. M. Frieze, and M. Mitzenmacher, "Min-wise independent permutations", *Journal of Computer and System Sciences*, vol. 60, no. 3, 2000, pp. 630–659. DOI: 10.1006/jcss.1999.1690.

[303] P. Indyk and R. Motwani, "Approximate nearest neighbors: Towards removing the curse of dimensionality", in *Thirtieth Annual ACM Symposium on the Theory of Computing (STOC)*, J. S. Vitter, Ed., pp. 604–613, ACM, 1998. DOI: 10.1145/276698.276876.

[304] A. Gionis, P. Indyk, and R. Motwani, "Similarity search in high dimensions via hashing", in *Conference on Very Large Databases (VLDB)*, M. P. Atkinson, M. E. Orlowska, P. Valduriez, S. B. Zdonik, and M. L. Brodie, Eds., pp. 518–529, Morgan Kaufmann, 1999. [Online]. Available: http://www.vldb.org/conf/1999/P49.pdf.

[305] J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of Massive Datasets, 3rd Edition*. Cambridge University Press, 2020. [Online]. Available: http://www.mmds.org/.

[306] S. Chaudhuri, K. Ganjam, V. Ganti, and R. Motwani, "Robust and efficient fuzzy match for online data cleaning", in *ACM Conference on Management of Data (SIGMOD)*, 2003. DOI: 10.1145/872757.872796.

[307] W. Cohen, P. Ravikumar, and S. Fienberg, "A comparison of string metrics for matching names and records", in *KDD Workshop on Data Cleaning and Object Consolidation*, pp. 73–78, 2003.

[308] S. Das, P. S. G. C., A. Doan, J. F. Naughton, G. Krishnan, R. Deep, E. Arcaute, V. Raghavendra, and Y. Park, "Falcon: Scaling up hands-off crowdsourced entity matching to build cloud services", in *ACM Conference on Management of Data (SIGMOD)*, pp. 1431–1446, ACM, 2017. DOI: 10.1145/3035918.3035960.

[309] E. Rahm and P. A. Bernstein, "A survey of approaches to automatic schema matching", *Journal of the VLDB*, vol. 10, no. 4, 2001, pp. 334–350. DOI: 10.1007/s007780100057.

Full text available at: http://dx.doi.org/10.1561/1900000064

[310] E. Rahm and E. Peukert, "Large-scale schema matching", in *Encyclopedia of Big Data Technologies*, Springer, 2019. DOI: 10.1007/978-3-319-63962-8_330-1.

[311] D. Firmani, S. Galhotra, B. Saha, and D. Srivastava, "Robust entity resolution using a crowdoracle", *IEEE Data Engineering Bulletin*, vol. 41, no. 2, 2018, pp. 91–103. [Online]. Available: http://sites.computer.org/debull/A18june/p91.pdf.

[312] J. Sun, D. Deng, I. F. Ilyas, G. Li, S. Madden, M. Ouzzani, M. Stonebraker, and N. Tang, "Technical report: Optimizing human involvement for entity matching and consolidation", *CoRR*, vol. abs/1906.06574, 2019. arXiv: 1906.06574. [Online]. Available: http://arxiv.org/abs/1906.06574.

[313] E. Rahm and E. Peukert, "Large scale entity resolution", in *Encyclopedia of Big Data Technologies*, Springer, 2019. DOI: 10.1007/978-3-319-63962-8_4-1.

[314] A. Saeedi, E. Peukert, and E. Rahm, "Comparative evaluation of distributed clustering schemes for multi-source entity resolution", in *Advances in Databases and Information Systems - 21st European Conference, ADBIS*, ser. Lecture Notes in Computer Science, vol. 10509, pp. 278–293, Springer, 2017. DOI: 10.1007/978-3-319-66917-5_19.

[315] Y. Li, J. Li, Y. Suhara, A. Doan, and W. Tan, "Deep entity matching with pre-trained language models", *Proceedings of the VLDB Endowment*, vol. 14, no. 1, 2020, pp. 50–60. DOI: 10.14778/3421424.3421431.

[316] W. Zhang, H. Wei, B. Sisman, X. L. Dong, C. Faloutsos, and D. Page, "Autoblock: A hands-off blocking framework for entity matching", in *ACM Conference on Web Search and Data Mining (WSDM)*, pp. 744–752, ACM, 2020. DOI: 10.1145/3336191.3371813.

[317] P. Li, X. Cheng, X. Chu, Y. He, and S. Chaudhuri, "Auto-fuzzyjoin: Auto-program fuzzy similarity joins without labeled examples", in *ACM Conference on Management of Data (SIGMOD)*, 2021.

Full text available at: http://dx.doi.org/10.1561/1900000064

[318] S. E. Whang, D. Menestrina, G. Koutrika, M. Theobald, and H. Garcia-Molina, "Entity resolution with iterative blocking", in *ACM Conference on Management of Data (SIGMOD)*, 2009. DOI: 10.1145/1559845.1559870.

[319] A. Madaan, A. Mittal, Mausam, G. Ramakrishnan, and S. Sarawagi, "Numerical relation extraction with minimal supervision", in *Conference on Artificial Intelligence (AAAI)*, 2016. [Online]. Available: http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12486.

[320] S. Roy, T. Vieira, and D. Roth, "Reasoning about quantities in natural language", *Transactions of the Association for Computational Linguistics (TACL)*, vol. 3, 2015, pp. 1–13. [Online]. Available: https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/452.

[321] O. Alonso and T. Sellam, "Quantitative information extraction from social data", in *ACM Conference on Research and Development in Information Retrieval (SIGIR)*, 2018. DOI: 10.1145/3209978.3210133.

[322] A. Waagmeester, G. Stupp, S. Burgstaller-Muehlbacher, B. M. Good, M. Griffith, O. L. Griffith, K. Hanspers, H. Hermjakob, T. S. Hudson, K. Hybiske, *et al.*, "Wikidata as a knowledge graph for the life sciences", *ELife*, vol. 9, 2020.

[323] F. M. Suchanek, G. Kasneci, and G. Weikum, " Yago - A Large Ontology from Wikipedia and WordNet ", in *Journal of Web Semantics (JWS)*, 2008.

[324] A. Sahuguet and F. Azavant, "Building light-weight wrappers for legacy web data-sources using W4F", in *Conference on Very Large Databases (VLDB)*, 1999. [Online]. Available: http://www.vldb.org/conf/1999/P72.pdf.

[325] E. Ferrara, P. D. Meo, G. Fiumara, and R. Baumgartner, "Web data extraction, applications and techniques: A survey", *Knowledge Based Systems*, vol. 70, 2014, pp. 301–323. DOI: 10.1016/j.knosys.2014.07.007.

Full text available at: http://dx.doi.org/10.1561/1900000064

[326] M. F. Hanafi, A. Abouzied, L. Chiticariu, and Y. Li, "SEER: auto-generating information extraction rules from user-specified examples", in *ACM Conference on Human Factors in Computing Systems (CHI)*, 2017. DOI: 10.1145/3025453.3025540.

[327] N. Kushmerick, D. S. Weld, and R. B. Doorenbos, "Wrapper induction for information extraction", in *Joint Conference on Artificial Intelligence (IJCAI)*, 1997.

[328] S. Soderland, "Learning information extraction rules for semi-structured and free text", *Machine Learning*, vol. 34, no. 1-3, 1999, pp. 233–272. DOI: 10.1023/A:1007562322031.

[329] N. Kushmerick, "Wrapper induction: Efficiency and expressiveness", *Artificial Intelligence (AI)*, vol. 118, no. 1-2, 2000, pp. 15–68. DOI: 10.1016/S0004-3702(99)00100-9.

[330] I. Muslea, S. Minton, and C. A. Knoblock, "Hierarchical wrapper induction for semistructured information sources", *Autonomous Agents and Multi-Agent Systems*, vol. 4, no. 1/2, 2001, pp. 93–114. DOI: 10.1023/A:1010022931168.

[331] R. Baumgartner, S. Flesca, and G. Gottlob, "Visual web information extraction with lixto", in *Conference on Very Large Databases (VLDB)*, 2001. [Online]. Available: http://www.vldb.org/conf/2001/P119.pdf.

[332] F. Reiss, S. Raghavan, R. Krishnamurthy, H. Zhu, and S. Vaithyanathan, "An algebraic approach to rule-based information extraction", in *IEEE International Conference on Data Engineering (ICDE)*, 2008. DOI: 10.1109/ICDE.2008.4497502.

[333] L. Chiticariu, R. Krishnamurthy, Y. Li, S. Raghavan, F. Reiss, and S. Vaithyanathan, "Systemt: An algebraic approach to declarative information extraction", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2010. [Online]. Available: https://www.aclweb.org/anthology/P10-1014/.

[334] W. Shen, A. Doan, J. F. Naughton, and R. Ramakrishnan, "Declarative information extraction using datalog with embedded extraction predicates", in *Conference on Very Large Databases (VLDB)*, 2007. [Online]. Available: http://www.vldb.org/conf/2007/papers/research/p1033-shen.pdf.

Full text available at: http://dx.doi.org/10.1561/1900000064

[335] W. Shen, P. DeRose, R. McCann, A. Doan, and R. Ramakrishnan, "Toward best-effort information extraction", in *ACM Conference on Management of Data (SIGMOD)*, 2008. DOI: 10.1145/1376616.1376718.

[336] P. Bohannon, S. Merugu, C. Yu, V. Agarwal, P. DeRose, A. S. Iyer, A. Jain, V. Kakade, M. Muralidharan, R. Ramakrishnan, and W. Shen, "Purple SOX extraction management system", *SIGMOD Record*, vol. 37, no. 4, 2008, pp. 21–27. DOI: 10.1145/1519103.1519107.

[337] A. Jain, P. G. Ipeirotis, and L. Gravano, "Building query optimizers for information extraction: The sqout project", *SIGMOD Record*, vol. 37, no. 4, 2008, pp. 28–34. DOI: 10.1145/1519103.1519108.

[338] P. G. Ipeirotis, E. Agichtein, P. Jain, and L. Gravano, "Towards a query optimizer for text-centric tasks", *ACM Transactions on Database Systems (TODS)*, vol. 32, no. 4, 2007, p. 21. DOI: 10.1145/1292609.1292611.

[339] P. Pantel and M. Pennacchiotti, "Espresso: Leveraging generic patterns for automatically harvesting semantic relations", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, The Association for Computer Linguistics, 2006. DOI: 10.3115/1220175.1220190.

[340] F. M. Suchanek, G. Ifrim, and G. Weikum, "Combining linguistic and statistical analysis to extract relations from web documents", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2006. DOI: 10.1145/1150402.1150492.

[341] M. Mintz, S. Bills, R. Snow, and D. Jurafsky, "Distant supervision for relation extraction without labeled data", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2009. [Online]. Available: https://www.aclweb.org/anthology/P09-1113/.

[342] F. M. Suchanek, M. Sozio, and G. Weikum, "Sofie: A self-organizing framework for information extraction", in *The Web Conference (WWW)*, 2009.

Full text available at: http://dx.doi.org/10.1561/1900000064

[343] N. Nakashole, M. Theobald, and G. Weikum, "Scalable knowledge harvesting with high precision and high recall", in *ACM Conference on Web Search and Data Mining (WSDM)*, 2011.

[344] R. Hoffmann, C. Zhang, and D. S. Weld, "Learning 5000 relational extractors", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2010. [Online]. Available: https://www.aclweb.org/anthology/P10-1030/.

[345] M. Surdeanu, D. McClosky, J. Tibshirani, J. Bauer, A. X. Chang, V. I. Spitkovsky, and C. D. Manning, "A simple distant supervision approach for the TAC-KBP slot filling task", in *Text Analysis Conference (TAC)*, 2010. [Online]. Available: https://tac.nist.gov/publications/2010/participant.papers/Stanford.proceedings.pdf.

[346] H. Ji, R. Grishman, H. T. Dang, K. Griffitt, and J. Ellis, "Overview of the tac 2010 knowledge base population track", in *Text Analysis Conference (TAC)*, 2010.

[347] L. Yao, A. Haghighi, S. Riedel, and A. McCallum, "Structured relation discovery using generative models", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2011.

[348] D. Roth and W.-t. Yih, "Global inference for entity and relation identification via a linear programming formulation", *Introduction to statistical relational learning*, 2007, pp. 553–580.

[349] S. Riedel, L. Yao, and A. McCallum, "Modeling relations and their mentions without labeled text", in *European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD)*, 2010. DOI: 10.1007/978-3-642-15939-8_10.

[350] L. Yao, S. Riedel, and A. McCallum, "Collective cross-document relation extraction without labelled data", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2010. [Online]. Available: https://www.aclweb.org/anthology/D10-1099/.

[351] J. Zhu, Z. Nie, J. Wen, B. Zhang, and W. Ma, "Simultaneous record detection and attribute labeling in web data extraction", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2006. DOI: 10.1145/1150402.1150457.

Full text available at: http://dx.doi.org/10.1561/1900000064

[352] S. Zheng, R. Song, J. Wen, and C. L. Giles, "Efficient record-level wrapper induction", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2009. DOI: 10.1145/1645953.1645962.

[353] J. Zhu, Z. Nie, X. Liu, B. Zhang, and J. Wen, "Statsnowball: A statistical approach to extracting entity relationships", in *The Web Conference (WWW)*, 2009. DOI: 10.1145/1526709.1526724.

[354] Q. Li and H. Ji, "Incremental joint extraction of entity mentions and relations", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2014. DOI: 10.3115/v1/p14-1038.

[355] X. Ren, Z. Wu, W. He, M. Qu, C. R. Voss, H. Ji, T. F. Abdelzaher, and J. Han, "Cotype: Joint extraction of typed entities and relations with knowledge bases", in *The Web Conference (WWW)*, 2017. DOI: 10.1145/3038912.3052708.

[356] M. Qu, X. Ren, Y. Zhang, and J. Han, "Weakly-supervised relation extraction by pattern-enhanced embedding learning", in *The Web Conference (WWW)*, 2018. DOI: 10.1145/3178876.3186024.

[357] A. Culotta and J. S. Sorensen, "Dependency tree kernels for relation extraction", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2004. [Online]. Available: https://www.aclweb.org/anthology/P04-1054/.

[358] R. C. Bunescu and R. J. Mooney, "Subsequence kernels for relation extraction", in *Neural Information Processing Systems (NeurIPS)*, 2005. [Online]. Available: http://papers.nips.cc/paper/2787-subsequence-kernels-for-relation-extraction.

[359] A. Moschitti, "A study on convolution kernels for shallow statistic parsing", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2004. [Online]. Available: https://www.aclweb.org/anthology/P04-1043/.

[360] A. Moschitti, "Efficient convolution kernels for dependency and constituent syntactic trees", in *European Conference on Machine Learning (ECML)*, 2006. DOI: 10.1007/11871842_32.

Full text available at: http://dx.doi.org/10.1561/1900000064

[361] G. Zhou, M. Zhang, D. Ji, and Q. Zhu, "Tree kernel-based relation extraction with context-sensitive structured parse tree information", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2007. [Online]. Available: https://www.aclweb.org/anthology/D07-1076/.

[362] D. Tikk, P. Thomas, P. Palaga, J. Hakenberg, and U. Leser, "A comprehensive benchmark of kernel methods to extract protein-protein interactions from literature", *PLoS Computational Biology*, vol. 6, no. 7, 2010. DOI: 10.1371/journal.pcbi.1000837.

[363] H. Liu, L. Hunter, V. Keselj, and K. Verspoor, "Approximate subgraph matching-based literature mining for biomedical events and relations", *PloS one*, vol. 8, no. 4, 2013.

[364] D. Gildea and D. Jurafsky, "Automatic labeling of semantic roles", *Computational Linguistics*, vol. 28, no. 3, 2002, pp. 245–288. DOI: 10.1162/089120102760275983.

[365] M. Palmer, D. Gildea, and N. Xue, *Semantic Role Labeling*. Morgan & Claypool Publishers, 2010. DOI: 10.2200/S00239ED1V01Y200912HLT006.

[366] V. Punyakanok, D. Roth, and W. Yih, "The importance of syntactic parsing and inference in semantic role labeling", *Computational Linguistics*, vol. 34, no. 2, 2008, pp. 257–287. DOI: 10.1162/coli.2008.34.2.257.

[367] L. He, K. Lee, M. Lewis, and L. Zettlemoyer, "Deep semantic role labeling: What works and what's next", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2017. DOI: 10.18653/v1/P17-1044.

[368] N. FitzGerald, J. Michael, L. He, and L. Zettlemoyer, "Large-scale QA-SRL parsing", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2018. DOI: 10.18653/v1/P18-1191.

[369] O. Etzioni, M. Banko, and M. J. Cafarella, "Machine reading", in *Conference on Artificial Intelligence (AAAI)*, 2006. [Online]. Available: http://www.aaai.org/Library/AAAI/2006/aaai06-239.php.

Full text available at: http://dx.doi.org/10.1561/1900000064

[370] S. Chakrabarti, M. van den Berg, and B. Dom, "Focused crawling: A new approach to topic-specific web resource discovery", *Computer Networks*, vol. 31, no. 11-16, 1999, pp. 1623–1640. DOI: 10.1016/S1389-1286(99)00052-3.

[371] S. Chakrabarti, *Mining the web - discovering knowledge from hypertext data*. Morgan Kaufmann, 2003.

[372] F. Menczer, G. Pant, and P. Srinivasan, "Topical web crawlers: Evaluating adaptive algorithms", *ACM Transactions on Internet Technology (TOIT)*, vol. 4, no. 4, 2004, pp. 378–419. DOI: 10.1145/1031114.1031117.

[373] S. Sarawagi and V. G. V. Vydiswaran, "Learning to extract information from large domain-specific websites using sequential models", *SIGKDD Explorations*, vol. 6, no. 2, 2004, pp. 61–66. DOI: 10.1145/1046456.1046464.

[374] E. C. Dragut, W. Meng, and C. T. Yu, *Deep Web Query Interface Understanding and Integration*. Morgan & Claypool Publishers, 2012. DOI: 10.2200/S00419ED1V01Y201205DTM026.

[375] S. Sizov, M. Theobald, S. Siersdorfer, G. Weikum, J. Graupmann, M. Biwer, and P. Zimmer, "The bingo! system for information portal generation and expert web search", in *Conference on Innovative Data Systems Research (CIDR)*, 2003. [Online]. Available: http://www-db.cs.wisc.edu/cidr/cidr2003/program/p7.pdf.

[376] L. Barbosa and J. Freire, "An adaptive crawler for locating hidden web entry points", in *The Web Conference (WWW)*, 2007. DOI: 10.1145/1242572.1242632.

[377] K. Vieira, L. Barbosa, A. S. da Silva, J. Freire, and E. S. de Moura, "Finding seeds to bootstrap focused crawlers", *World Wide Web*, vol. 19, no. 3, 2016, pp. 449–474. DOI: 10.1007/s11280-015-0331-7.

[378] X. Wang, X. L. Dong, Y. Li, and A. Meliou, "Midas: Finding the right web sources to fill knowledge gaps", in *ACM Conference on Management of Data (SIGMOD)*, 2019.

[379] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, "The eigentrust algorithm for reputation management in P2P networks", in *The Web Conference (WWW)*, 2003. DOI: 10.1145/775152.775242.

Full text available at: http://dx.doi.org/10.1561/1900000064

[380] S. Pandey and C. Olston, "User-centric web crawling", in *The Web Conference (WWW)*, pp. 401–411, ACM, 2005. DOI: 10.1145/1060745.1060805.

[381] D. Cai, S. Yu, J. Wen, and W. Ma, "Block-based web search", in *ACM Conference on Research and Development in Information Retrieval (SIGIR)*, 2004. DOI: 10.1145/1008992.1009070.

[382] J. Zhu, B. Zhang, Z. Nie, J. Wen, and H. Hon, "Webpage understanding: An integrated approach", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2007. DOI: 10.1145/1281192.1281288.

[383] S. Zheng, R. Song, J. Wen, and D. Wu, "Joint optimization of wrapper generation and template detection", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2007. DOI: 10.1145/1281192.1281287.

[384] P. Gulhane, A. Madaan, R. R. Mehta, J. Ramamirtham, R. Rastogi, S. Satpal, S. H. Sengamedu, A. Tengli, and C. Tiwari, "Web-scale information extraction with vertex", in *IEEE International Conference on Data Engineering (ICDE)*, 2011. DOI: 10.1109/ICDE.2011.5767842.

[385] Q. Hao, R. Cai, Y. Pang, and L. Zhang, "From one tree to a forest: A unified solution for structured web data extraction", in *ACM Conference on Research and Development in Information Retrieval (SIGIR)*, 2011. DOI: 10.1145/2009916.2010020.

[386] C. Lockard, X. L. Dong, P. Shiralkar, and A. Einolghozati, "CERES: distantly supervised relation extraction from the semi-structured web", *Proceedings of the VLDB Endowment*, vol. 11, no. 10, 2018, pp. 1084–1096. DOI: 10.14778/3231751.3231758.

[387] M. J. Cafarella, A. Y. Halevy, H. Lee, J. Madhavan, C. Yu, D. Z. Wang, and E. Wu, "Ten years of webtables", *Proceedings of the VLDB Endowment*, vol. 11, no. 12, 2018, pp. 2140–2149. DOI: 10.14778/3229863.3240492.

[388] Y. Oulabi and C. Bizer, "Extending cross-domain knowledge bases with long tail entities using web table data", in *Conference on Extending Database Technology (EDBT)*, 2019. DOI: 10.5441/002/edbt.2019.34.

Full text available at: http://dx.doi.org/10.1561/1900000064

[389] B. Kruit, P. A. Boncz, and J. Urbani, "Extracting novel facts from tables for knowledge graph completion", in *International Semantic Web Conference (ISWC)*, ser. Lecture Notes in Computer Science, vol. 11778, pp. 364–381, Springer, 2019. DOI: 10.1007/978-3-030-30793-6_21.

[390] D. Wang, P. Shiralkar, C. Lockard, B. Huang, X. L. Dong, and M. Jiang, "Tcn: Table convolutional network for web table interpretation", in *The Web Conference (WWW)*, 2021.

[391] V. T. Ho, K. Pal, S. Razniewski, K. Berberich, and G. Weikum, "Extracting contextualized quantity facts from web tables", in *The Web Conference (WWW)*, 2021.

[392] S. Krause, H. Li, H. Uszkoreit, and F. Xu, "Large-scale learning of relation-extraction rules with distant supervision from the web", in *International Semantic Web Conference (ISWC)*, 2012. DOI: 10.1007/978-3-642-35176-1_17.

[393] S. Krause, L. Hennig, A. Moro, D. Weissenborn, F. Xu, H. Uszkoreit, and R. Navigli, "Sar-graphs: A language resource connecting linguistic knowledge with semantic relations from knowledge graphs", *Journal of Web Semantics (JWS)*, vol. 37-38, 2016, pp. 112–131. DOI: 10.1016/j.websem.2016.03.004.

[394] P. Ernst, A. Siu, and G. Weikum, "Highlife: Higher-arity fact harvesting", in *The Web Conference (WWW)*, 2018. DOI: 10.1145/3178876.3186000.

[395] O. Lehmberg and C. Bizer, "Synthesizing n-ary relations from web tables", in *International Conference on Web Intelligence, Mining and Semantics*, 2019. DOI: 10.1145/3326467.3326480.

[396] B. Kruit, P. A. Boncz, and J. Urbani, "Extracting n-ary facts from wikipedia table clusters", in *ACM Conference on Information and Knowledge Management (CIKM)*, pp. 655–664, ACM, 2020. DOI: 10.1145/3340531.3412027.

[397] C. Bizer, K. Eckert, R. Meusel, H. Mühleisen, M. Schuhmacher, and J. Völker, "Deployment of rdfa, microdata, and microformats on the web - A quantitative analysis", in *International Semantic Web Conference (ISWC)*, 2013. DOI: 10.1007/978-3-642-41338-4_2.

Full text available at: http://dx.doi.org/10.1561/1900000064

[398] R. Yu, U. Gadiraju, B. Fetahu, O. Lehmberg, D. Ritze, and S. Dietze, "Knowmore - knowledge base augmentation with structured web markup", *Semantic Web Journal (SWJ)*, vol. 10, no. 1, 2019, pp. 159–180. DOI: 10.3233/SW-180304.

[399] R. Socher, B. Huval, C. D. Manning, and A. Y. Ng, "Semantic compositionality through recursive matrix-vector spaces", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2012. [Online]. Available: https://www.aclweb.org/anthology/D12-1110/.

[400] K. Hashimoto, M. Miwa, Y. Tsuruoka, and T. Chikayama, "Simple customization of recursive neural networks for semantic relation classification", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2013. [Online]. Available: https://www.aclweb.org/anthology/D13-1137/.

[401] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, "Relation classification via convolutional deep neural network", in *Conference on Computational Linguistics (COLING)*, 2014. [Online]. Available: https://www.aclweb.org/anthology/C14-1220/.

[402] Y. Liu, F. Wei, S. Li, H. Ji, M. Zhou, and H. Wang, "A dependency-based neural network for relation classification", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2015. DOI: 10.3115/v1/p15-2047.

[403] M. R. Gormley, M. Yu, and M. Dredze, "Improved relation extraction with feature-rich compositional embedding models", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2015. DOI: 10.18653/v1/d15-1205.

[404] Y. Xu, L. Mou, G. Li, Y. Chen, H. Peng, and Z. Jin, "Classifying relations via long short term memory networks along shortest dependency paths", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2015. DOI: 10.18653/v1/d15-1206.

[405] C. N. dos Santos, B. Xiang, and B. Zhou, "Classifying relations by ranking with convolutional neural networks", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2015. DOI: 10.3115/v1/p15-1061.

Full text available at: http://dx.doi.org/10.1561/1900000064

[406] Y. Lin, S. Shen, Z. Liu, H. Luan, and M. Sun, "Neural relation extraction with selective attention over instances", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2016. DOI: 10.18653/v1/p16-1200.

[407] G. Ji, K. Liu, S. He, and J. Zhao, "Distant supervision for relation extraction with sentence-level attention and entity descriptions", in *Conference on Artificial Intelligence (AAAI)*, 2017. [Online]. Available: http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14491.

[408] Y. Zhang, V. Zhong, D. Chen, G. Angeli, and C. D. Manning, "Position-aware attention and supervised data improve slot filling", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2017. DOI: 10.18653/v1/d17-1004.

[409] D. Sorokin and I. Gurevych, "Context-aware representations for knowledge base relation extraction", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2017. DOI: 10.18653/v1/d17-1188.

[410] C. Christodoulopoulos and A. Mittal, "Simple large-scale relation extraction from unstructured text", in *Conference on Language Resources and Evaluation (LREC)*, 2018. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2018/summaries/7.html.

[411] Y. Yao, D. Ye, P. Li, X. Han, Y. Lin, Z. Liu, Z. Liu, L. Huang, J. Zhou, and M. Sun, "Docred: A large-scale document-level relation extraction dataset", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019. DOI: 10.18653/v1/p19-1074.

[412] L. B. Soares, N. FitzGerald, J. Ling, and T. Kwiatkowski, "Matching the blanks: Distributional similarity for relation learning", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019. DOI: 10.18653/v1/p19-1279.

[413] M. Miwa and M. Bansal, "End-to-end relation extraction using lstms on sequences and tree structures", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2016. DOI: 10.18653/v1/p16-1105.

Full text available at: http://dx.doi.org/10.1561/1900000064

[414] T. Bansal, P. Verga, N. Choudhary, and A. McCallum, "Simultaneously linking entities and extracting relations from biomedical text without mention-level supervision", *CoRR*, vol. abs/1912.01070, 2019. [Online]. Available: http://arxiv.org/abs/1912.01070.

[415] Y. Luan, L. He, M. Ostendorf, and H. Hajishirzi, "Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018. DOI: 10.18653/v1/d18-1360.

[416] N. Peng, H. Poon, C. Quirk, K. Toutanova, and W. Yih, "Cross-sentence n-ary relation extraction with graph lstms", *Transactions of the Association for Computational Linguistics (TACL)*, vol. 5, 2017, pp. 101–115. [Online]. Available: https://transacl.org/ojs/index.php/tacl/article/view/1028.

[417] S. K. Sahu, F. Christopoulou, M. Miwa, and S. Ananiadou, "Inter-sentence relation extraction with document-level graph convolutional neural network", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019. DOI: 10.18653/v1/p19-1423.

[418] B. D. Trisedya, G. Weikum, J. Qi, and R. Zhang, "Neural relation extraction for knowledge base enrichment", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019. DOI: 10.18653/v1/p19-1023.

[419] D. Zeng, K. Liu, Y. Chen, and J. Zhao, "Distant supervision for relation extraction via piecewise convolutional neural networks", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2015. DOI: 10.18653/v1/d15-1203.

[420] Y. Yuan, L. Liu, S. Tang, Z. Zhang, Y. Zhuang, S. Pu, F. Wu, and X. Ren, "Cross-relation cross-bag attention for distantly-supervised relation extraction", in *Conference on Artificial Intelligence (AAAI)*, 2019. DOI: 10.1609/aaai.v33i01.3301419.

[421] G. Zheng, S. Mukherjee, X. L. Dong, and F. Li, "Opentag: Open attribute value extraction from product profiles", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2018. DOI: 10.1145/3219819.3219839.

Full text available at: http://dx.doi.org/10.1561/1900000064

[422] Y. Wu, D. Bamman, and S. J. Russell, "Adversarial training for relation extraction", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2017. DOI: 10.18653/v1/d17-1187.

[423] T. Sun, C. Zhang, Y. Ji, and Z. Hu, "Reinforcement learning for distantly supervised relation extraction", *IEEE Access*, vol. 7, 2019, pp. 98 023–98 033. DOI: 10.1109/ACCESS.2019.2930340.

[424] R. Takanobu, T. Zhang, J. Liu, and M. Huang, "A hierarchical framework for relation extraction with reinforcement learning", in *Conference on Artificial Intelligence (AAAI)*, 2019. DOI: 10.1609/aaai.v33i01.33017072.

[425] X. Han, T. Gao, Y. Lin, H. Peng, Y. Yang, C. Xiao, Z. Liu, P. Li, M. Sun, and J. Zhou, "More data, more relations, more context and more openness: A review and outlook for relation extraction", *CoRR*, vol. abs/2004.03186, 2020.

[426] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need", in *Neural Information Processing Systems (NeurIPS)*, 2017. [Online]. Available: http://papers.nips.cc/paper/7181-attention-is-all-you-need.

[427] Harvard NLP Group, "The annotated transformer", *https://nlp.seas.harvard.edu/2018/04/03/attention.html*, 2018.

[428] J. Alammar, "The illustrated transformer", *http://jalammar.github.io/illustrated-transformer*, 2018.

[429] P. Verga, E. Strubell, and A. McCallum, "Simultaneously self-attending to all mentions for full-abstract biological relation extraction", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2018. DOI: 10.18653/v1/n18-1080.

[430] C. Alt, M. Hübner, and L. Hennig, "Fine-tuning pre-trained transformer language models to distantly supervised relation extraction", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019. DOI: 10.18653/v1/p19-1134.

Full text available at: http://dx.doi.org/10.1561/1900000064

[431] H. Wang, M. Tan, M. Yu, S. Chang, D. Wang, K. Xu, X. Guo, and S. Potdar, "Extracting multiple-relations in one-pass with pre-trained transformers", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019. DOI: 10.18653/v1/p19-1132.

[432] S. Wu and Y. He, "Enriching pre-trained language model with entity information for relation classification", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2019. DOI: 10.1145/3357384.3358119.

[433] Z. Bouraoui, J. Camacho-Collados, and S. Schockaert, "Inducing relational knowledge from BERT", *CoRR*, vol. abs/1911.12753, 2019.

[434] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "Squad: 100, 000+ questions for machine comprehension of text", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2016. DOI: 10.18653/v1/d16-1264.

[435] O. Levy, M. Seo, E. Choi, and L. Zettlemoyer, "Zero-shot relation extraction via reading comprehension", in *Conference on Computational Natural Language Learning (CoNLL)*, 2017. DOI: 10.18653/v1/K17-1034.

[436] X. Li, F. Yin, Z. Sun, X. Li, A. Yuan, D. Chai, M. Zhou, and J. Li, "Entity-relation extraction as multi-turn question answering", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019. DOI: 10.18653/v1/p19-1129.

[437] R. Das, T. Munkhdalai, X. Yuan, A. Trischler, and A. McCallum, "Building dynamic knowledge graphs from text using machine reading comprehension", in *International Conference on Learning Representations (ICLR)*, 2019. [Online]. Available: https://openreview.net/forum?id=S1lhbnRqF7.

[438] C. Lockard, P. Shiralkar, and X. L. Dong, "Openceres: When open information extraction meets the semi-structured web", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2019. DOI: 10.18653/v1/n19-1309.

Full text available at: http://dx.doi.org/10.1561/1900000064

[439] S. Saha and Mausam, "Open information extraction from conjunctive sentences", in *Conference on Computational Linguistics (COLING)*, pp. 2288–2299, Association for Computational Linguistics, 2018. [Online]. Available: https://www.aclweb.org/anthology/C18-1194/.

[440] M. Banko, M. J. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni, "Open information extraction from the web", in *Joint Conference on Artificial Intelligence (IJCAI)*, 2007. [Online]. Available: http://ijcai.org/Proceedings/07/Papers/429.pdf.

[441] F. Wu and D. S. Weld, "Open information extraction using wikipedia", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2010. [Online]. Available: https://www.aclweb.org/anthology/P10-1013/.

[442] A. Fader, S. Soderland, and O. Etzioni, "Identifying relations for open information extraction", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2011.

[443] Mausam, M. Schmitz, S. Soderland, R. Bart, and O. Etzioni, "Open language learning for information extraction", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2012. [Online]. Available: https://www.aclweb.org/anthology/D12-1048/.

[444] O. Etzioni, A. Fader, J. Christensen, S. Soderland, and Mausam, "Open information extraction: The second generation", in *Joint Conference on Artificial Intelligence (IJCAI)*, 2011. DOI: 10.5591/978-1-57735-516-8/IJCAI11-012.

[445] L. D. Corro and R. Gemulla, "Clausie: Clause-based open information extraction", in *The Web Conference (WWW)*, 2013. DOI: 10.1145/2488388.2488420.

[446] Mausam, "Open information extraction systems and downstream applications", in *Joint Conference on Artificial Intelligence (IJCAI)*, 2016. [Online]. Available: http://www.ijcai.org/Abstract/16/604.

[447] G. Angeli, M. J. J. Premkumar, and C. D. Manning, "Leveraging linguistic structure for open domain information extraction", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2015. DOI: 10.3115/v1/p15-1034.

Full text available at: http://dx.doi.org/10.1561/1900000064

[448] M. Yahya, S. Whang, R. Gupta, and A. Halevy, "Renoun: Fact extraction for nominal attributes", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014.

[449] H. Pal and Mausam, "Demonyms and compound relational nouns in nominal open IE", in *Workshop on Automated Knowledge Base Construction (AKBC)*, 2016. DOI: 10.18653/v1/w16-1307.

[450] K. Gashteovski, R. Gemulla, and L. D. Corro, "Minie: Minimizing facts in open information extraction", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2017. DOI: 10.18653/v1/d17-1278.

[451] N. Bhutani, H. V. Jagadish, and D. R. Radev, "Nested propositions in open information extraction", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2016. DOI: 10.18653/v1/d16-1006.

[452] R. E. Prasojo, M. Kacimi, and W. Nutt, "Stuffie: Semantic tagging of unlabeled facets using fine-grained information extraction", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2018. DOI: 10.1145/3269206.3271812.

[453] M. Palmer, P. R. Kingsbury, and D. Gildea, "The proposition bank: An annotated corpus of semantic roles", *Computational Linguistics*, vol. 31, no. 1, 2005, pp. 71–106. DOI: 10.1162/089120105363026 4.

[454] C. J. Fillmore, C. Wooters, and C. F. Baker, "Building a large lexical databank which provides deep semantics", in *Pacific Asia Conference on Language, Information and Computation (PACLIC)*, 2001. [Online]. Available: http://hdl.handle.net/2065/12202.

[455] S. S. Pradhan, E. H. Hovy, M. P. Marcus, M. Palmer, L. A. Ramshaw, and R. M. Weischedel, "Ontonotes: A unified relational semantic representation", *International Journal of Semantic Computing*, vol. 1, no. 4, 2007, pp. 405–419. DOI: 10.1142/S1793351X07000251.

[456] K. Gashteovski, R. Gemulla, B. Kotnis, S. Hertling, and C. Meilicke, "On aligning openie extractions with knowledge bases: A case study", in *Workshop on Evaluation and Comparison of NLP Systems, co-located with EMNLP 2020*, pp. 143–154, 2020.

Full text available at: http://dx.doi.org/10.1561/1900000064

[457] G. Stanovsky, J. Michael, L. Zettlemoyer, and I. Dagan, "Supervised open information extraction", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2018. DOI: 10.18653/v1/n18-1081.

[458] L. He, M. Lewis, and L. Zettlemoyer, "Question-answer driven semantic role labeling: Using natural language to annotate natural language", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2015. DOI: 10.18653/v1/d15-1076.

[459] J. Michael, G. Stanovsky, L. He, I. Dagan, and L. Zettlemoyer, "Crowdsourcing question-answer meaning representations", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2018. DOI: 10.18653/v1/n18-2089.

[460] G. Liu, X. Li, J. Wang, M. Sun, and P. Li, "Extracting knowledge from web text with monte carlo tree search", in *The Web Conference (WWW)*, 2020. DOI: 10.1145/3366423.3380010.

[461] K. Kolluru, V. Adlakha, S. Aggarwal, Mausam, and S. Chakrabarti, "Openie6: Iterative grid labeling and coordination analysis for open information extraction", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 3748–3761, Association for Computational Linguistics, 2020. DOI: 10.18653/v1/2020.emnlp-main.306.

[462] R. Gupta, A. Y. Halevy, X. Wang, S. E. Whang, and F. Wu, "Biperpedia: An ontology for search applications", *Proceedings of the VLDB Endowment*, vol. 7, no. 7, 2014, pp. 505–516. DOI: 10.14778/2732286.2732288.

[463] M. Pasca and B. V. Durme, "What you seek is what you get: Extraction of class attributes from query logs", in *Joint Conference on Artificial Intelligence (IJCAI)*, 2007. [Online]. Available: http://ijcai.org/Proceedings/07/Papers/455.pdf.

[464] M. Pasca, "The role of query sessions in interpreting compound noun phrases", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2015. DOI: 10.1145/2806416.2806571.

[465] M. Bronzi, V. Crescenzi, P. Merialdo, and P. Papotti, "Extraction and integration of partially overlapping web sources", *Proceedings of the VLDB Endowment*, vol. 6, no. 10, 2013, pp. 805–816. DOI: 10.14778/2536206.2536209.

Full text available at: http://dx.doi.org/10.1561/1900000064

[466] X. Zhu and Z. Ghahramani, "Learning from labeled and unlabeled data with label propagation", *Technical Report, Carnegie Mellon University*, 2002.

[467] C. Lockard, P. Shiralkar, X. L. Dong, and H. Hajishizi, "Zeroshotceres: Zero-shot relation extraction from semi-structured webpages", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020.

[468] M. Yakout, K. Ganjam, K. Chakrabarti, and S. Chaudhuri, "Infogather: Entity augmentation and attribute discovery by holistic matching with web tables", in *ACM Conference on Management of Data (SIGMOD)*, 2012. DOI: 10.1145/2213836.2213848.

[469] M. Zhang and K. Chakrabarti, "Infogather+: Semantic matching and annotation of numeric and time-varying attributes in web tables", in *ACM Conference on Management of Data (SIGMOD)*, 2013. DOI: 10.1145/2463676.2465276.

[470] Y. Ibrahim, M. Riedewald, G. Weikum, and D. Zeinalipour-Yazti, "Bridging quantities in tables and text", in *IEEE International Conference on Data Engineering (ICDE)*, 2019. DOI: 10.1109/ICDE.2019.00094.

[471] N. Lao, T. Mitchell, and W. W. Cohen, "Random walk inference and learning in a large scale knowledge base", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2011.

[472] N. Lao, A. Subramanya, F. Pereira, and W. W. Cohen, "Reading the web with learned syntactic-semantic inference rules", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2012.

[473] D. Lin and P. Pantel, "DIRT – discovery of inference rules from text", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2001. [Online]. Available: http://portal.acm.org/citation.cfm?id=502512.502559.

[474] T. Chklovski and P. Pantel, "Verbocean: Mining the web for fine-grained semantic verb relations", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2004. [Online]. Available: https://www.aclweb.org/anthology/W04-3205/.

Full text available at: http://dx.doi.org/10.1561/1900000064

[475] N. Nakashole, G. Weikum, and F. M. Suchanek, "PATTY: A taxonomy of relational patterns with semantic types", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2012. [Online]. Available: https://www.aclweb.org/anthology/D12-1104/.

[476] C. D. Bovi, L. Telesca, and R. Navigli, "Large-scale information extraction from textual definitions through deep syntactic and semantic analysis", *Transactions of the Association for Computational Linguistics (TACL)*, vol. 3, 2015, pp. 529–543. [Online]. Available: https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/660.

[477] R. Barzilay and K. R. McKeown, "Extracting paraphrases from a parallel corpus", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2001. DOI: 10.3115/1073012.1073020.

[478] C. J. Bannard and C. Callison-Burch, "Paraphrasing with bilingual parallel corpora", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2005. DOI: 10.3115/1219840.1219914.

[479] J. Ganitkevitch, B. V. Durme, and C. Callison-Burch, "PPDB: the paraphrase database", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2013. [Online]. Available: https://www.aclweb.org/anthology/N13-1092/.

[480] E. Pavlick, P. Rastogi, J. Ganitkevitch, B. V. Durme, and C. Callison-Burch, "PPDB 2.0: Better paraphrase ranking, fine-grained entailment relations, word embeddings, and style classification", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2015. DOI: 10.3115/v1/p15-2070.

[481] A. Grycner and G. Weikum, "POLY: mining relational paraphrases from multilingual sentences", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2016. DOI: 10.18653/v1/d16-1236.

[482] M. Faruqui and S. Kumar, "Multilingual open relation extraction using cross-lingual projection", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2015. DOI: 10.3115/v1/n15-1151.

Full text available at: http://dx.doi.org/10.1561/1900000064

[483] Y. He, K. Chakrabarti, T. Cheng, and T. Tylenda, "Automatic discovery of attribute synonyms using query logs and table corpora", in *The Web Conference (WWW)*, 2016. DOI: 10.1145/2872427.2874816.

[484] A. Yates and O. Etzioni, "Unsupervised methods for determining object and relation synonyms on the web", *Journal of Artificial Intelligence Research*, vol. 34, 2009, pp. 255–296. DOI: 10.1613/jair.2772.

[485] B. Min, S. Shi, R. Grishman, and C. Lin, "Ensemble semantics for large-scale unsupervised relation extraction", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2012. [Online]. Available: https://www.aclweb.org/anthology/D12-1094/.

[486] J. Pujara, H. Miao, L. Getoor, and W. W. Cohen, "Knowledge graph identification", in *International Semantic Web Conference (ISWC)*, 2013. DOI: 10.1007/978-3-642-41335-3_34.

[487] L. Galárraga, G. Heitz, K. Murphy, and F. M. Suchanek, "Canonicalizing open knowledge bases", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2014. DOI: 10.1145/2661829.2662073.

[488] K. Pal, V. T. Ho, and G. Weikum, "Co-clustering triples from open information extraction", in *CoDS-COMAD Conference*, 2020. DOI: 10.1145/3371158.3371183.

[489] D. B. Nguyen, A. Abujabal, K. Tran, M. Theobald, and G. Weikum, "Query-driven on-the-fly knowledge base construction", *Proceedings of the VLDB Endowment*, vol. 11, no. 1, 2017, pp. 66–79. DOI: 10.14778/3151113.3151119.

[490] X. Lin, H. Li, H. Xin, Z. Li, and L. Chen, "Kbpearl: A knowledge base population system supported by joint entity and relation linking", *Proceedings of the VLDB Endowment*, vol. 13, no. 7, 2020, pp. 1035–1049. [Online]. Available: http://www.vldb.org/pvldb/vol13/p1035-lin.pdf.

[491] Y. Koren, R. M. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems", *IEEE Computer*, vol. 42, no. 8, 2009, pp. 30–37. DOI: 10.1109/MC.2009.263.

Full text available at: http://dx.doi.org/10.1561/1900000064

[492] Y. Koren and R. M. Bell, "Advances in collaborative filtering", in *Recommender Systems Handbook*, 2015, pp. 77–118. DOI: 10.1007/978-1-4899-7637-6_3.

[493] Y. Zhang and X. Chen, "Explainable recommendation: A survey and new perspectives", *Foundations and Trends in Information Retrieval*, vol. 14, no. 1, 2020, pp. 1–101. DOI: 10.1561/1500000066.

[494] S. Riedel, L. Yao, A. McCallum, and B. M. Marlin, "Relation extraction with matrix factorization and universal schemas", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2013. [Online]. Available: https://www.aclweb.org/anthology/N13-1008/.

[495] L. Yao, S. Riedel, and A. McCallum, "Universal schema for entity type prediction", in *Workshop on Automated Knowledge Base Construction (AKBC)*, 2013. DOI: 10.1145/2509558.2509572.

[496] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, "A review of relational machine learning for knowledge graphs", *Proceedings of the IEEE*, vol. 104, no. 1, 2016, pp. 11–33. DOI: 10.1109/JPROC.2015.2483592.

[497] S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu, "A survey on knowledge graphs: Representation, acquisition and applications", *CoRR*, vol. abs/2002.00388, 2020.

[498] D. Ruffinelli, S. Broscheit, and R. Gemulla, "You CAN teach an old dog new tricks! on training knowledge graph embeddings", in *International Conference on Learning Representations (ICLR)*, 2020. [Online]. Available: https://openreview.net/forum?id=BkxSmlBFvr.

[499] P. Verga, A. Neelakantan, and A. McCallum, "Generalizing to unseen entities and entity pairs with row-less universal schema", in *Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 2017. DOI: 10.18653/v1/e17-1058.

Full text available at: http://dx.doi.org/10.1561/1900000064

[500] P. Jain, S. Murty, Mausam, and S. Chakrabarti, "Mitigating the effect of out-of-vocabulary entity pairs in matrix factorization for KB inference", in *Joint Conference on Artificial Intelligence (IJCAI)*, pp. 4122–4129, ijcai.org, 2018. DOI: 10.24963/ijcai.2018/573.

[501] D. Erdös and P. Miettinen, "Discovering facts with boolean tensor tucker decomposition", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2013. DOI: 10.1145/2505515.2507846.

[502] K. Chang, W. Yih, B. Yang, and C. Meek, "Typed tensor decomposition of knowledge bases for relation extraction", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014. DOI: 10.3115/v1/d14-1165.

[503] P. Verga, D. Belanger, E. Strubell, B. Roth, and A. McCallum, "Multilingual relation extraction using compositional universal schema", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2016. DOI: 10.18653/v1/n16-1103.

[504] D. Zhang, S. Mukherjee, C. Lockard, L. Dong, and A. McCallum, "Openki: Integrating open information extraction and knowledge bases with relation inference", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2019. [Online]. Available: https://aclweb.org/anthology/papers/N/N19/N19-1083/.

[505] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms", in *International Conference on Machine Learning (ICML)*, 2006. DOI: 10.1145/1143844.1143865.

[506] D. Koller and N. Friedman, *Probabilistic Graphical Models - Principles and Techniques*. MIT Press, 2009. [Online]. Available: http://mitpress.mit.edu/catalog/item/default.asp?ttype=2%5C&tid=11886.

[507] H. Paulheim, "Knowledge graph refinement: A survey of approaches and evaluation methods", *Semantic Web Journal (SWJ)*, 2017.

Full text available at: http://dx.doi.org/10.1561/1900000064

[508] M. Färber, F. Bartscherer, C. Menne, and A. Rettinger, "Linked data quality of dbpedia, freebase, opencyc, wikidata, and YAGO", *Semantic Web Journal (SWJ)*, vol. 9, no. 1, 2018, pp. 77–129. DOI: 10.3233/SW-170275.

[509] A. Piscopo and E. Simperl, "What we talk about when we talk about wikidata quality: A literature survey", in *International Symposium on Open Collaboration*, 2019. DOI: 10.1145/3306446.3340822.

[510] N. Heist, S. Hertling, D. Ringler, and H. Paulheim, "Knowledge graphs on the web - an overview", in *Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges*, IOS Press, 2020, pp. 3–22. DOI: 10.3233/SSW200009.

[511] G. Bordea, E. Lefever, and P. Buitelaar, "Semeval-2016 task 13: Taxonomy extraction evaluation", in *Workshop on Semantic Evaluation (SemEval)*, 2016.

[512] J. Gao, X. Li, Y. E. Xu, B. Sisman, X. L. Dong, and J. Yang, "Efficient knowledge graph accuracy evaluation", *Proceedings of the VLDB Endowment*, vol. 12, no. 11, 2019, pp. 1679–1691. DOI: 10.14778/3342263.3342642.

[513] A. Sabharwal and H. Sedghi, "How good are my predictions? efficiently approximating precision-recall curves for massive datasets", in *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2017.

[514] C. H. Tan, E. Agichtein, P. Ipeirotis, and E. Gabrilovich, "Trust, but verify: Predicting contribution quality for knowledge base construction and curation", in *ACM Conference on Web Search and Data Mining (WSDM)*, 2014. DOI: 10.1145/2556195.2556227.

[515] A. Doan, R. Ramakrishnan, and A. Y. Halevy, "Crowdsourcing systems on the world-wide web", *Commun. ACM*, vol. 54, no. 4, 2011, pp. 86–96. DOI: 10.1145/1924421.1924442.

[516] A. I. Chittilappilly, L. Chen, and S. Amer-Yahia, "A survey of general-purpose crowdsourcing techniques", *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 28, no. 9, 2016, pp. 2246–2266. DOI: 10.1109/TKDE.2016.2555805.

[517] O. Alonso, *The Practice of Crowdsourcing*. Morgan & Claypool Publishers, 2019. DOI: 10.2200/S00904ED1V01Y201903ICR066.

Full text available at: http://dx.doi.org/10.1561/1900000064

[518] S. Heindorf, M. Potthast, B. Stein, and G. Engels, "Vandalism detection in wikidata", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2016.

[519] M. D. Lieberman and J. J. Lin, "You are where you edit: Locating wikipedia contributors through edit histories", in *International Conference on Weblogs and Social Media (ICWSM)*, 2009. [Online]. Available: http://aaai.org/ocs/index.php/ICWSM/09/paper/view/204.

[520] S. Kumar, R. West, and J. Leskovec, "Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes", in *The Web Conference (WWW)*, 2016. DOI: 10.1145/2872427.2883085.

[521] Y. Hua, C. Danescu-Niculescu-Mizil, D. Taraborelli, N. Thain, J. Sorensen, and L. Dixon, "Wikiconv: A corpus of the complete conversational history of a large online collaborative community", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018. DOI: 10.18653/v1/d18-1305.

[522] X. L. Dong, E. Gabrilovich, K. Murphy, V. Dang, W. Horn, C. Lugaresi, S. Sun, and W. Zhang, "Knowledge-based trust: Estimating the trustworthiness of web sources", *Proceedings of the VLDB Endowment*, vol. 8, no. 9, 2015, pp. 938–949. DOI: 10.14778/2777598.2777603.

[523] F. Li, X. L. Dong, A. Langen, and Y. Li, "Knowledge verification for longtail verticals", *Proceedings of the VLDB Endowment*, vol. 10, no. 11, 2017, pp. 1370–1381. DOI: 10.14778/3137628.3137646.

[524] X. Yin, J. Han, and P. S. Yu, "Truth discovery with multiple conflicting information providers on the web", *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 20, no. 6, 2008, pp. 796–808. DOI: 10.1109/TKDE.2007.190745.

[525] X. Li, W. Meng, and C. T. Yu, "T-verifier: Verifying truthfulness of fact statements", in *IEEE International Conference on Data Engineering (ICDE)*, 2011. DOI: 10.1109/ICDE.2011.5767859.

Full text available at: http://dx.doi.org/10.1561/1900000064

[526] N. Nakashole and T. M. Mitchell, "Language-aware truth assessment of fact candidates", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2014. DOI: 10.3115/v1/p14-1095.

[527] M. H. Gad-Elrab, D. Stepanova, J. Urbani, and G. Weikum, "Exfakt: A framework for explaining facts over knowledge graphs and text", in *ACM Conference on Web Search and Data Mining (WSDM)*, 2019. DOI: 10.1145/3289600.3290996.

[528] Y. Li, J. Gao, C. Meng, Q. Li, L. Su, B. Zhao, W. Fan, and J. Han, "A survey on truth discovery", *SIGKDD Explorations*, vol. 17, no. 2, 2015, pp. 1–16. DOI: 10.1145/2897350.2897352.

[529] B. Trushkowsky, T. Kraska, M. J. Franklin, P. Sarkar, and V. Ramachandran, "Crowdsourcing enumeration queries: Estimators and interfaces", *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 27, no. 7, 2015, pp. 1796–1809. DOI: 10.1109/TKDE.2014.2339857.

[530] M. Luggen, D. Difallah, C. Sarasua, G. Demartini, and P. Cudré-Mauroux, "Non-parametric class completeness estimators for collaborative knowledge graphs - the case of wikidata", in *International Semantic Web Conference (ISWC)*, 2019.

[531] A. Soulet, A. Giacometti, B. Markhoff, and F. M. Suchanek, "Representativeness of knowledge bases with the generalized benford's law", in *International Semantic Web Conference (ISWC)*, 2018. DOI: 10.1007/978-3-030-00671-6_22.

[532] A. Hopkinson, A. Gurdasani, D. Palfrey, and A. Mittal, "Demand-weighted completeness prediction for a knowledge base", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2018. DOI: 10.18653/v1/n18-3025.

[533] S. Razniewski and P. Das, "Structured knowledge: Have we made progress? an extrinsic study of kb coverage over 19 years", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2020.

[534] P. Mirza, S. Razniewski, F. Darari, and G. Weikum, "Cardinal virtues: Extracting relation cardinalities from text", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2017. DOI: 10.18653/v1/P17-2055.

Full text available at: http://dx.doi.org/10.1561/1900000064

[535] P. Mirza, S. Razniewski, F. Darari, and G. Weikum, "Enriching knowledge bases with counting quantifiers", in *International Semantic Web Conference (ISWC)*, 2018. DOI: 10.1007/978-3-030-00671-6_11.

[536] S. Razniewski, N. Jain, P. Mirza, and G. Weikum, "Coverage of information extraction from sentences and paragraphs", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019.

[537] S. Razniewski, F. M. Suchanek, and W. Nutt, "But what do we actually know?", in *Workshop on Automated Knowledge Base Construction (AKBC)*, 2016. DOI: 10.18653/v1/w16-1308.

[538] L. A. Galárraga, C. Teflioudi, K. Hose, and F. M. Suchanek, *AMIE: association rule mining under incomplete evidence in ontological knowledge bases*, 2013. DOI: 10.1145/2488388.2488425.

[539] L. Galárraga, S. Razniewski, A. Amarilli, and F. M. Suchanek, "Predicting completeness in knowledge bases", in *ACM Conference on Web Search and Data Mining (WSDM)*, 2017. DOI: 10.1145/3018661.3018739.

[540] A. Motro, "Integrity= validity+ completeness", *ACM Transactions on Database Systems (TODS)*, vol. 14, no. 4, 1989, pp. 480–502.

[541] S. Razniewski, F. Korn, W. Nutt, and D. Srivastava, "Identifying the extent of completeness of query answers over partially complete databases", in *ACM Conference on Management of Data (SIGMOD)*, 2015. DOI: 10.1145/2723372.2750544.

[542] F. Darari, W. Nutt, G. Pirrò, and S. Razniewski, "Completeness management for RDF data sources", *ACM Transactions on the Web (TWEB)*, vol. 12, no. 3, 2018, 18:1–18:53. DOI: 10.1145/3196248.

[543] W. Lang, R. V. Nehme, E. Robinson, and J. F. Naughton, "Partial results in database systems", in *ACM Conference on Management of Data (SIGMOD)*, 2014.

Full text available at: http://dx.doi.org/10.1561/1900000064

[544] S. Ghosh, S. Razniewski, and G. Weikum, "Uncovering hidden semantics of set information in knowledge bases", *Journal of Web Semantics (JWS)*, vol. 64, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1570826820300317.

[545] V. Balaraman, S. Razniewski, and W. Nutt, "Recoin: Relative completeness in wikidata", in *Companion of the The Web Conference*, 2018. DOI: 10.1145/3184558.3191641.

[546] H. Arnaout, S. Razniewski, and G. Weikum, "Enriching knowledge bases with interesting negative statements", in *Conference on Automatic Knowledge Base Construction (AKBC)*, 2020.

[547] P. E. O'Neil and E. J. O'Neil, *Database: Principles, Programming, and Performance, Second Edition*. Morgan Kaufmann, 2000.

[548] H. Knublauch and D. Kontokostas, "Shapes constraint language (SHACL), https://www.w3.org/TR/2017/REC-shacl-20170720/", W3C Recommendation, 2017.

[549] J. E. L. Gayo, E. Prud'hommeaux, H. R. Solbrig, and J. M. A. Rodriguez, "Validating and describing linked data portals using rdf shape expressions.", in *Workshop on Linked Data Quality (LDQ), co-located with International Conference on Semantic Systems (SEMANTICS)*, 2014.

[550] I. Boneva, J. E. L. Gayo, and E. G. Prud'hommeaux, "Semantics and validation of shapes schemas for RDF", in *International Semantic Web Conference (ISWC)*, 2017. DOI: 10.1007/978-3-319-68288-4_7.

[551] F. Baader, I. Horrocks, C. Lutz, and U. Sattler, *An Introduction to Description Logic*. Cambridge University Press, 2017. [Online]. Available: http://www.cambridge.org/de/academic/subjects/computer-science/knowledge-management-databases-and-data-mining/introduction-description-logic?format=PB%5C#17zVGeWD2TZUeu6s.97.

[552] L. E. Bertossi, *Database Repairing and Consistent Query Answering*. Morgan & Claypool Publishers, 2011. DOI: 10.2200/S00379ED1V01Y201108DTM020.

Full text available at: http://dx.doi.org/10.1561/1900000064

[553] L. E. Bertossi, "Database repairs and consistent query answering: Origins and further developments", in *ACM Symposium on Principles of Database Systems (PODS)*, 2019. DOI: 10.1145/3294052.3322190.

[554] M. Bienvenu, "A short survey on inconsistency handling in ontology-mediated query answering", *KI-Künstliche Intelligenz*, 2020, pp. 1–9.

[555] I. F. Ilyas, V. Markl, P. J. Haas, P. Brown, and A. Aboulnaga, "CORDS: automatic discovery of correlations and soft functional dependencies", in *ACM Conference on Management of Data (SIGMOD)*, 2004. DOI: 10.1145/1007568.1007641.

[556] F. Chiang and R. J. Miller, "Discovering data quality rules", *Proceedings of the VLDB Endowment*, vol. 1, no. 1, 2008, pp. 1166–1177. DOI: 10.14778/1453856.1453980.

[557] H. Saxena, L. Golab, and I. F. Ilyas, "Distributed implementations of dependency discovery algorithms", *Proceedings of the VLDB Endowment*, vol. 12, no. 11, 2019, pp. 1624–1636. DOI: 10.14778/3342263.3342638.

[558] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases", in *Conference on Very Large Databases (VLDB)*, 1994. [Online]. Available: http://www.vldb.org/conf/1994/P487.PDF.

[559] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. I. Verkamo, "Fast discovery of association rules", in *Advances in Knowledge Discovery and Data Mining*, pp. 307–328, AAAI/MIT Press, 1996.

[560] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques, 3rd edition*. Morgan Kaufmann, 2011. [Online]. Available: http://hanj.cs.illinois.edu/bk3/.

[561] S. Muggleton and L. D. Raedt, "Inductive logic programming: Theory and methods", *Journal of Logic Programming*, vol. 19/20, 1994, pp. 629–679. DOI: 10.1016/0743-1066(94)90035-3.

[562] Q. Zeng, J. M. Patel, and D. Page, "Quickfoil: Scalable inductive logic programming", *Proceedings of the VLDB Endowment*, vol. 8, no. 3, 2014, pp. 197–208. DOI: 10.14778/2735508.2735510.

Full text available at: http://dx.doi.org/10.1561/1900000064

[563] A. Cropper, S. Dumancic, and S. H. Muggleton, "Turning 30: New ideas in inductive logic programming", *CoRR*, vol. abs/2002.11002, 2020.

[564] L. Galárraga, C. Teflioudi, K. Hose, and F. M. Suchanek, "Fast rule mining in ontological knowledge bases with AMIE+", *Journal of the VLDB*, vol. 24, no. 6, 2015, pp. 707–730. DOI: 10.1007/s00778-015-0394-1.

[565] J. Lajus, L. Galárraga, and F. M. Suchanek, "Fast and exact rule mining with AMIE 3", in *European Semantic Web Conference (ESWC)*, 2020. DOI: 10.1007/978-3-030-49461-2_3.

[566] S. Ortona, V. V. Meduri, and P. Papotti, "Rudik: Rule discovery in knowledge bases", *Proceedings of the VLDB Endowment*, vol. 11, no. 12, 2018, pp. 1946–1949. DOI: 10.14778/3229863.3236231.

[567] S. Ortona, V. V. Meduri, and P. Papotti, "Robust discovery of positive and negative rules in knowledge bases", in *IEEE International Conference on Data Engineering (ICDE)*, 2018. DOI: 10.1109/ICDE.2018.00108.

[568] N. Lao, E. Minkov, and W. W. Cohen, "Learning relational features with backward random walks", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2015. DOI: 10.3115/v1/p15-1065.

[569] T. P. Tanon, D. Stepanova, S. Razniewski, P. Mirza, and G. Weikum, "Completeness-aware rule learning from knowledge graphs", in *Joint Conference on Artificial Intelligence (IJCAI)*, 2018.

[570] C. Meilicke, M. W. Chekol, D. Ruffinelli, and H. Stuckenschmidt, "Anytime bottom-up rule learning for knowledge graph completion", in *Joint Conference on Artificial Intelligence (IJCAI)*, 2019. DOI: 10.24963/ijcai.2019/435.

[571] M. H. Gad-Elrab, D. Stepanova, J. Urbani, and G. Weikum, "Exception-enriched rule learning from knowledge graphs", in *International Semantic Web Conference (ISWC)*, 2016.

[572] J. Völker and M. Niepert, "Statistical schema induction", in *European Semantic Web Conference (ESWC)*, 2011.

Full text available at: http://dx.doi.org/10.1561/1900000064

[573] B. Minaei-Bidgoli, R. Barmaki, and M. Nasiri, "Mining numerical association rules via multi-objective genetic algorithms", *Information Sciences*, 2013.

[574] N. Heist and H. Paulheim, "Information extraction from co-occurring similar entities", in *The Web Conference (WWW)*, 2021.

[575] D. Stepanova, M. H. Gad-Elrab, and V. T. Ho, "Rule induction and reasoning over knowledge graphs", in *Reasoning Web Summer School*, pp. 142–172, Springer, 2018. DOI: 10.1007/978-3-030-00338-8_6.

[576] M. Nickel, V. Tresp, and H. Kriegel, "Factorizing YAGO: scalable machine learning for linked data", in *The Web Conference (WWW)*, 2012. DOI: 10.1145/2187836.2187874.

[577] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data", in *Neural Information Processing Systems (NeurIPS)*, 2013. [Online]. Available: http://papers.nips.cc/paper/5071-translating-embeddings-for-modeling-multi-relational-data.

[578] R. Socher, D. Chen, C. D. Manning, and A. Y. Ng, "Reasoning with neural tensor networks for knowledge base completion", in *Neural Information Processing Systems (NeurIPS)*, 2013. [Online]. Available: http://papers.nips.cc/paper/5028-reasoning-with-neural-tensor-networks-for-knowledge-base-completion.

[579] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion", in *Conference on Artificial Intelligence (AAAI)*, 2015. [Online]. Available: http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9571.

[580] R. Zhang, B. D. Trisedya, M. Li, Y. Jiang, and J. Qi, "A comprehensive survey on knowledge graph entity alignment via representation learning", *CoRR*, vol. abs/2103.15059, 2021. arXiv: 2103.15059. [Online]. Available: https://arxiv.org/abs/2103.15059.

Full text available at: http://dx.doi.org/10.1561/1900000064

[581] T. Rekatsinas, X. Chu, I. F. Ilyas, and C. Ré, "Holoclean: Holistic data repairs with probabilistic inference", *Proceedings of the VLDB Endowment*, vol. 10, no. 11, 2017, pp. 1190–1201. DOI: 10.14778/3137628.3137631.

[582] M. Arenas, L. E. Bertossi, and J. Chomicki, "Consistent query answers in inconsistent databases", in *ACM Symposium on Principles of Database Systems (PODS)*, 1999. DOI: 10.1145/303976.303983.

[583] P. Bohannon, M. Flaster, W. Fan, and R. Rastogi, "A cost-based model and effective heuristic for repairing constraints by value modification", in *ACM Conference on Management of Data (SIGMOD)*, 2005. DOI: 10.1145/1066157.1066175.

[584] X. Chu, I. F. Ilyas, and P. Papotti, "Holistic data cleaning: Putting violations into context", in *IEEE International Conference on Data Engineering (ICDE)*, 2013. DOI: 10.1109/ICDE.2013.6544847.

[585] N. Prokoshyna, J. Szlichta, F. Chiang, R. J. Miller, and D. Srivastava, "Combining quantitative and logical data cleaning", *Proceedings of the VLDB Endowment*, vol. 9, no. 4, 2015, pp. 300–311. DOI: 10.14778/2856318.2856325.

[586] X. L. Dong and F. Naumann, "Data fusion - resolving data conflicts for integration", *Proceedings of the VLDB Endowment*, vol. 2, no. 2, 2009, pp. 1654–1655. DOI: 10.14778/1687553.1687620.

[587] X. L. Dong, L. Berti-Équille, and D. Srivastava, "Data fusion: Resolving conflicts from multiple sources", in *Handbook of Data Quality, Research and Practice*, 2013, pp. 293–318. DOI: 10.1007/978-3-642-36257-6_13.

[588] X. L. Dong, E. Gabrilovich, G. Heitz, W. Horn, K. Murphy, S. Sun, and W. Zhang, "From data fusion to knowledge fusion", *Proceedings of the VLDB Endowment*, vol. 7, no. 10, 2014, pp. 881–892. DOI: 10.14778/2732951.2732962.

[589] *Handbook of Satisfiability*, vol. 185, Frontiers in Artificial Intelligence and Applications, IOS Press, 2009.

Full text available at: http://dx.doi.org/10.1561/1900000064

[590] F. Rossi, P. van Beek, and T. Walsh, *Handbook of Constraint Programming*, vol. 2. Elsevier, 2006. [Online]. Available: http://www.sciencedirect.com/science/bookseries/15746526/2.

[591] C. M. Li and F. Manyà, "Maxsat, hard and soft constraints", in *Handbook of Satisfiability*, IOS Press, 2009, pp. 613–631. DOI: 10.3233/978-1-58603-929-5-613.

[592] A. Schrijver, *Theory of linear and integer programming*, ser. Wiley-Interscience series in discrete mathematics and optimization. Wiley, 1999.

[593] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge University Press, 1995. DOI: 10.1017/cbo9780511814075.

[594] V. V. Vazirani, *Approximation algorithms*. Springer, 2001. [Online]. Available: http://www.springer.com/computer/theoretical+computer+science/book/978-3-540-65367-7.

[595] P. M. Domingos and D. Lowd, *Markov Logic: An Interface Layer for Artificial Intelligence*. Morgan & Claypool Publishers, 2009. DOI: 10.2200/S00206ED1V01Y200907AIM007.

[596] D. Roth, "On the hardness of approximate reasoning", *Artificial Intelligence (AI)*, vol. 82, no. 1-2, 1996, pp. 273–302. DOI: 10.1016/0004-3702(94)00092-1.

[597] S. Riedel, "Improving the accuracy and efficiency of MAP inference for markov logic", *CoRR*, vol. abs/1206.3282, 2012.

[598] M. Richardson and P. M. Domingos, "Markov logic networks", *Machine Learning*, vol. 62, no. 1-2, 2006, pp. 107–136. DOI: 10.1007/s10994-006-5833-1.

[599] H. Poon and P. M. Domingos, "Unsupervised ontology induction from text", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2010. [Online]. Available: https://www.aclweb.org/anthology/P10-1031/.

[600] C. Zhang, C. Ré, M. J. Cafarella, J. Shin, F. Wang, and S. Wu, "Deepdive: Declarative knowledge base construction", *Communication of the ACM*, vol. 60, no. 5, 2017, pp. 93–102. DOI: 10.1145/3060586.

Full text available at: http://dx.doi.org/10.1561/1900000064

[601] C. Zhang, J. Shin, C. Ré, M. J. Cafarella, and F. Niu, "Extracting databases from dark data with deepdive", in *ACM Conference on Management of Data (SIGMOD)*, 2016. DOI: 10.1145/2882903.2904442.

[602] M. Kejriwal and P. Szekely, "Knowledge graphs for social good: An entity-centric search engine for the human trafficking domain", *IEEE Transactions on Big Data*, 2017.

[603] F. Niu, C. Ré, A. Doan, and J. W. Shavlik, "Tuffy: Scaling up statistical inference in markov logic networks using an RDBMS", *Proceedings of the VLDB Endowment*, vol. 4, no. 6, 2011, pp. 373–384. DOI: 10.14778/1978665.1978669.

[604] S. H. Bach, M. Broecheler, B. Huang, and L. Getoor, "Hinge-loss markov random fields and probabilistic soft logic", *Journal of Machine Learning Research (JMLR)*, vol. 18, 2017, 109:1–109:67. [Online]. Available: http://jmlr.org/papers/v18/15-631.html.

[605] P. Kouki, J. Pujara, C. Marcum, L. M. Koehly, and L. Getoor, "Collective entity resolution in familial networks", in *IEEE Conference on Data Mining (ICDM)*, 2017. DOI: 10.1109/ICDM.2017.32.

[606] A. Grycner, G. Weikum, J. Pujara, J. R. Foulds, and L. Getoor, "RELLY: inferring hypernym relationships between relational phrases", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2015. DOI: 10.18653/v1/d15-1113.

[607] J. Pujara, H. Miao, L. Getoor, and W. W. Cohen, "Large-scale knowledge graph identification using PSL", in *AAAI Fall Symposia*, 2013. [Online]. Available: http://www.aaai.org/ocs/index.php/FSS/FSS13/paper/view/7593.

[608] M. Chang, L. Ratinov, and D. Roth, "Structured learning with constrained conditional models", *Machine Learning*, vol. 88, no. 3, 2012, pp. 399–431. DOI: 10.1007/s10994-012-5296-5.

Full text available at: http://dx.doi.org/10.1561/1900000064

[609] T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, B. Yang, J. Betteridge, A. Carlson, B. D. Mishra, M. Gardner, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohamed, N. Nakashole, E. A. Platanios, A. Ritter, M. Samadi, B. Settles, R. C. Wang, D. Wijaya, A. Gupta, X. Chen, A. Saparov, M. Greaves, and J. Welling, "Never-ending learning", *Communication of the ACM*, vol. 61, no. 5, 2018, pp. 103–115.

[610] P. Buneman, S. Khanna, and W.-C. Tan, "Why and where: A characterization of data provenance", in *International Conference on Database Theory (ICDT)*, 2001. DOI: 10.1007/3-540-44503-X_20.

[611] P. Buneman and W.-C. Tan, "Data provenance: What next?", *SIGMOD Record*, vol. 47, no. 3, 2018, pp. 5–16. DOI: 10.1145/3316416.3316418.

[612] A. Piscopo, L. Kaffee, C. Phethean, and E. Simperl, "Provenance information in a collaborative knowledge graph: An evaluation of wikidata external references", in *International Semantic Web Conference (ISWC)*, 2017. DOI: 10.1007/978-3-319-68288-4_32.

[613] A. Piscopo, C. Phethean, and E. Simperl, "What makes a good collaborative knowledge graph: Group composition and quality in wikidata", in *International Conference on Social Informatics*, 2017. DOI: 10.1007/978-3-319-67217-5_19.

[614] O. Hartig, "Provenance information in the web of data.", in *Workshop on Linked Data on the Web (LDOW), co-located with the Web Conference*, 2009.

[615] S. Harris and A. Seaborne, "SPARQL 1.1 query language, https://www.w3.org/TR/2013/REC-sparql11-query-20130321/", W3C, W3C Recommendation, 2013.

[616] G. Carothers, "RDF 1.1 n-quads, https://www.w3.org/TR/2014/REC-n-quads-20140225/", W3C, W3C Recommendation, 2014.

[617] O. Hartig, "Foundations of RDF⋆ and SPARQL⋆ (an alternative approach to statement-level metadata in RDF)", in *Alberto Mendelzon Workshop (AMW)*, 2017. [Online]. Available: http://ceur-ws.org/Vol-1912/paper12.pdf.

Full text available at: http://dx.doi.org/10.1561/1900000064

[618] O. Hartig, "Rdf* and sparql*: An alternative approach to annotate statements in RDF", in *International Semantic Web Conference (ISWC)*, 2017.

[619] T. P. Tanon and F. Suchanek, "Querying the edit history of wikidata", in *European Semantic Web Conference (ESWC)*, 2019.

[620] M. Verhagen, I. Mani, R. Sauri, J. Littman, R. Knippen, S. B. Jang, A. Rumshisky, J. Phillips, and J. Pustejovsky, "Automating temporal annotation with TARSQI", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2005. [Online]. Available: https://www.aclweb.org/anthology/P05-3021/.

[621] E. Kuzey, J. Vreeken, and G. Weikum, "A fresh look on knowledge bases: Distilling named events from news", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2014. DOI: 10.1145/2661829.2661984.

[622] E. Kuzey and G. Weikum, "Extraction of temporal facts and events from wikipedia", in *Temporal Web Analytics Workshop, co-located with the Web Conference*, 2012. DOI: 10.1145/2169095.2169101.

[623] A. Sil and S. Cucerzan, "Towards temporal scoping of relational facts based on wikipedia data", in *Conference on Computational Natural Language Learning (CoNLL)*, 2014. DOI: 10.3115/v1/w14-1612.

[624] J. Ferguson, C. Lockard, D. S. Weld, and H. Hajishirzi, "Semi-supervised event extraction with paraphrase clusters", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, pp. 359–364, Association for Computational Linguistics, 2018. DOI: 10.18653/v1/n18-2058.

[625] D. Fernàndez-Cañellas, J. Espadaler, D. Rodriguez, B. Garolera, G. Canet, A. Colom, J. M. Rimmek, X. Giró-i-Nieto, E. Bou, and J. C. Riveiro, "Vlx-stories: Building an online event knowledge base with emerging entity detection", in *International Semantic Web Conference (ISWC)*, ser. Lecture Notes in Computer Science, vol. 11779, pp. 382–399, Springer, 2019. DOI: 10.1007/978-3-030-30796-7_24.

Full text available at: http://dx.doi.org/10.1561/1900000064

[626] S. Gottschalk and E. Demidova, "Eventkg - the hub of event knowledge on the web - and biographical timeline generation", *Semantic Web Journal (SWJ)*, vol. 10, no. 6, 2019, pp. 1039–1070. DOI: 10.3233/SW-190355.

[627] X. Wang, Z. Wang, X. Han, W. Jiang, R. Han, Z. Liu, J. Li, P. Li, Y. Lin, and J. Zhou, "MAVEN: A massive general domain event detection dataset", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1652–1671, Association for Computational Linguistics, 2020. DOI: 10.18653/v1/2020.emnlp-main.129.

[628] X. Ling and D. S. Weld, "Temporal information extraction", in *Conference on Artificial Intelligence (AAAI)*, 2010. [Online]. Available: http://www.aaai.org/ocs/index.php/AAAI/AAAI10/paper/view/1805.

[629] P. P. Talukdar, D. Wijaya, and T. M. Mitchell, "Coupled temporal scoping of relational facts", in *ACM Conference on Web Search and Data Mining (WSDM)*, 2012. DOI: 10.1145/2124295.2124307.

[630] P. P. Talukdar, D. Wijaya, and T. M. Mitchell, "Acquiring temporal constraints between relations", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2012. DOI: 10.1145/2396761.2396886.

[631] Y. Wang, B. Yang, L. Qu, M. Spaniol, and G. Weikum, "Harvesting facts from textual web sources by constrained label propagation", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2011. DOI: 10.1145/2063576.2063698.

[632] Y. Wang, M. Dylla, M. Spaniol, and G. Weikum, "Coupling label propagation and constraints for temporal fact extraction", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2012. [Online]. Available: https://www.aclweb.org/anthology/P12-2046/.

[633] N. UzZaman, H. Llorens, L. Derczynski, J. F. Allen, M. Verhagen, and J. Pustejovsky, "Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations", in *Workshop on Semantic Evaluation (SemEval)*, 2013. [Online]. Available: https://www.aclweb.org/anthology/S13-2001/.

Full text available at: http://dx.doi.org/10.1561/1900000064

[634] P. Jindal and D. Roth, "Extraction of events and temporal expressions from clinical narratives", *Journal of Biomedical Informatics*, vol. 46, no. 6, 2013, S13–S19. DOI: 10.1016/j.jbi.2013.08.010.

[635] P. Mirza and S. Tonelli, "CATENA: causal and temporal relation extraction from natural language texts", in *Conference on Computational Linguistics (COLING)*, 2016. [Online]. Available: https://www.aclweb.org/anthology/C16-1007/.

[636] S. S. Dasgupta, S. N. Ray, and P. P. Talukdar, "Hyte: Hyperplane-based temporally aware knowledge graph embedding", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 2001–2011, Association for Computational Linguistics, 2018. DOI: 10.18653/v1/d18-1225.

[637] Q. Ning, S. Subramanian, and D. Roth, "An improved neural baseline for temporal relation extraction", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019. DOI: 10.18653/v1/D19-1642.

[638] P. Jain, S. Rathi, Mausam, and S. Chakrabarti, "Temporal knowledge base completion: New algorithms and evaluation protocols", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 3733–3747, Association for Computational Linguistics, 2020. DOI: 10.18653/v1/2020.emnlp-main.305.

[639] M. Chen, H. Zhang, Q. Ning, M. Li, H. Ji, and D. Roth, "Event-centric natural language understanding (tutorial)", in *Conference on Artificial Intelligence (AAAI)*, 2021.

[640] J. Hoffart, Y. Altun, and G. Weikum, "Discovering emerging entities with ambiguous names", in *The Web Conference (WWW)*, 2014.

[641] J. Hoffart, D. Milchevski, G. Weikum, A. Anand, and J. Singh, "The knowledge awakens: Keeping knowledge bases fresh with emerging entities", in *Companion of the International Conference on World Wide Web*, 2016. DOI: 10.1145/2872518.2890537.

[642] P. Jansson and S. Liu, "Topic modelling enriched LSTM models for the detection of novel and emerging named entities from social media", in *International Conference on Big Data*, 2017. DOI: 10.1109/BigData.2017.8258462.

Full text available at: http://dx.doi.org/10.1561/1900000064

[643] J. Yeo, H. Cho, J. Park, and S. Hwang, "Multimodal KB harvesting for emerging spatial entities", *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 29, no. 5, 2017, pp. 1073–1086. DOI: 10.1109/TKDE.2017.2651805.

[644] S. Akasaki, N. Yoshinaga, and M. Toyoda, "Early discovery of emerging entities in microblogs", in *Joint Conference on Artificial Intelligence (IJCAI)*, 2019. DOI: 10.24963/ijcai.2019/678.

[645] S. Zhang, E. Meij, K. Balog, and R. Reinanda, "Novel entity discovery from web tables", in *The Web Conference (WWW)*, 2020. DOI: 10.1145/3366423.3380205.

[646] J. Chen, X. Chen, I. Horrocks, E. B. Myklebust, and E. Jiménez-Ruiz, "Correcting knowledge base assertions", in *The Web Conference (WWW)*, 2020. DOI: 10.1145/3366423.3380226.

[647] P.-H. Paris and F. M. Suchanek, "Non-named entities - the silent majority", in *European Semantic Web Conference (ESWC)*, 2021.

[648] J. W. Murdock, A. Kalyanpur, C. Welty, J. Fan, D. A. Ferrucci, D. Gondek, L. Zhang, and H. Kanayama, "Typing candidate answers using type coercion", *IBM Journal of Research and Development*, vol. 56, no. 3, 2012, p. 7. DOI: 10.1147/JRD.2012.2187036.

[649] G. Kasneci, M. Ramanath, F. M. Suchanek, and G. Weikum, "The YAGO-NAGA approach to knowledge discovery", *SIGMOD Record*, vol. 37, no. 4, 2008, pp. 41–47. DOI: 10.1145/1519103.1519110.

[650] J. Hoffart, F. M. Suchanek, K. Berberich, E. Lewis-Kelham, G. de Melo, and G. Weikum, " YAGO2: A Spatially and Temporally Enhanced Knowledge Base from Wikipedia ", in *The Web Conference (WWW)*, 2011.

[651] N. Karalis, G. M. Mandilaras, and M. Koubarakis, "Extending the YAGO2 knowledge graph with precise geospatial knowledge", in *International Semantic Web Conference (ISWC)*, pp. 181–197, 2019. DOI: 10.1007/978-3-030-30796-7_12.

[652] J. A. Biega, E. Kuzey, and F. M. Suchanek, " Inside YAGO2s: A Transparent Information Extraction Architecture ", in *The Web Conference (WWW)*, 2013.

Full text available at: http://dx.doi.org/10.1561/1900000064

[653] F. Mahdisoltani, J. A. Biega, and F. M. Suchanek, " YAGO3: A Knowledge Base from Multilingual Wikipedias ", in *Conference on Innovative Data Systems Research (CIDR)*, 2015.

[654] M. Ehrmann, F. Cecconi, D. Vannella, J. P. McCrae, P. Cimiano, and R. Navigli, "Representing multilingual data as linked data: The case of babelnet 2.0", in *Conference on Language Resources and Evaluation (LREC)*, 2014. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2014/summaries/810.html.

[655] T. Rebele, F. M. Suchanek, J. Hoffart, J. A. Biega, E. Kuzey, and G. Weikum, "YAGO: a multilingual knowledge base from Wikipedia, Wordnet, and Geonames ", in *International Semantic Web Conference (ISWC)*, 2016.

[656] T. P. Tanon, G. Weikum, and F. M. Suchanek, "YAGO 4: A Reason-able Knowledge Base ", in *European Semantic Web Conference (ESWC)*, 2020.

[657] S. Auer and J. Lehmann, "What have innsbruck and leipzig in common? extracting semantics from wiki content", in *European Semantic Web Conference (ESWC)*, 2007. DOI: 10.1007/978-3-540-72667-8_36.

[658] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, and C. Bizer, "Dbpedia - A large-scale, multilingual knowledge base extracted from wikipedia", *Semantic Web Journal (SWJ)*, vol. 6, no. 2, 2015, pp. 167–195. DOI: 10.3233/SW-140134.

[659] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann, "Dbpedia - A crystallization point for the web of data", *Journal of Web Semantics (JWS)*, vol. 7, no. 3, 2009, pp. 154–165. DOI: 10.1016/j.websem.2009.07.002.

[660] S. Hellmann, C. Stadler, J. Lehmann, and S. Auer, "Dbpedia live extraction", in *OTM Conferences*, 2009. DOI: 10.1007/978-3-642-05151-7_33.

[661] P. N. Mendes, M. Jakob, and C. Bizer, "Dbpedia: A multilingual cross-domain knowledge base", in *Conference on Language Resources and Evaluation (LREC)*, 2012.

Full text available at: http://dx.doi.org/10.1561/1900000064

[662] D. Kontokostas, C. Bratsas, S. Auer, S. Hellmann, I. Antoniou, and G. Metakides, "Internationalization of linked data: The case of the greek dbpedia edition", *Journal of Web Semantics (JWS)*, vol. 15, 2012, pp. 51–61. DOI: 10.1016/j.websem.2012.01.001.

[663] A. Ismayilov, D. Kontokostas, S. Auer, J. Lehmann, and S. Hellmann, "Wikidata through the eyes of dbpedia", *Semantic Web Journal (SWJ)*, vol. 9, no. 4, 2018, pp. 493–503. DOI: 10.3233/SW-170277.

[664] A. Carlson, J. Betteridge, R. C. Wang, E. R. H. Jr., and T. M. Mitchell, "Coupled semi-supervised learning for information extraction", in *ACM Conference on Web Search and Data Mining (WSDM)*, 2010. DOI: 10.1145/1718487.1718501.

[665] T. Mohamed, E. R. H. Jr., and T. M. Mitchell, "Discovering relations between noun categories", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2011. [Online]. Available: https://www.aclweb.org/anthology/D11-1134/.

[666] N. Lao, T. M. Mitchell, and W. W. Cohen, "Random walk inference and learning in A large scale knowledge base", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2011. [Online]. Available: https://www.aclweb.org/anthology/D11-1049/.

[667] S. Malyshev, M. Krötzsch, L. González, J. Gonsior, and A. Bielefeldt, "Getting the most out of wikidata: Semantic technology usage in wikipedia's knowledge graph", in *International Semantic Web Conference (ISWC)*, 2018. DOI: 10.1007/978-3-030-00668-6_23.

[668] M. Krötzsch, D. Vrandecic, and M. Völkel, "Semantic mediawiki", in *International Semantic Web Conference (ISWC)*, 2006. DOI: 10.1007/11926078_68.

[669] T. P. Tanon, D. Vrandecic, S. Schaffert, T. Steiner, and L. Pintscher, "From freebase to wikidata: The great migration", in *The Web Conference (WWW)*, 2016. DOI: 10.1145/2872427.2874809.

Full text available at: http://dx.doi.org/10.1561/1900000064

[670] J. Dittrich *et al.*, "Research report - use of wikidata in glam institutions, https://commons.wikimedia.org/wiki/File:Research_Report_\T1\textendash_Use_of_Wikidata_in_GLAM_institutions_(2019-11).pdf", *Wikimedia Deutschland e.V.*, 2019.

[671] J. Wu, K. M. Williams, H. Chen, M. Khabsa, C. Caragea, S. Tuarob, A. Ororbia, D. Jordan, P. Mitra, and C. L. Giles, "Citeseerx: AI in a digital library search engine", *AI Magazine*, vol. 36, no. 3, 2015, pp. 35–48. DOI: 10.1609/aimag.v36i3.2601.

[672] R. A. Al-Zaidy and C. L. Giles, "Extracting semantic relations for scholarly knowledge base construction", in *International Conference on Semantic Computing (ICSC)*, 2018. DOI: 10.1109/ICSC.2018.00017.

[673] W. Ammar, D. Groeneveld, C. Bhagavatula, I. Beltagy, M. Crawford, D. Downey, J. Dunkelberger, A. Elgohary, S. Feldman, V. Ha, *et al.*, "Construction of the literature graph in semantic scholar", *arXiv:1805.02262*, 2018.

[674] K. Lo, L. L. Wang, M. Neumann, R. Kinney, and D. S. Weld, "S2ORC: the semantic scholar open research corpus", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020. [Online]. Available: https://www.aclweb.org/anthology/2020.acl-main.447/.

[675] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, "Arnetminer: Extraction and mining of academic social networks", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 990–998, 2008. DOI: 10.1145/1401890.1402008.

[676] H. Wan, Y. Zhang, J. Zhang, and J. Tang, "Aminer: Search and mining of academic social networks", *Data Intelligence*, vol. 1, no. 1, 2019, pp. 58–76. DOI: 10.1162/dint_a_00006.

[677] A. Oelen, M. Y. Jaradeh, M. Stocker, and S. Auer, "Generate FAIR literature surveys with scholarly knowledge graphs", in *Joint Conference on Digital Libraries*, 2020. DOI: 10.1145/3383583.3398520.

Full text available at: http://dx.doi.org/10.1561/1900000064

[678] A. Brack, A. Hoppe, M. Stocker, S. Auer, and R. Ewerth, "Requirements analysis for an open research knowledge graph", in *International Conference on Theory and Practice of Digital Libraries (TPDL)*, 2020. DOI: 10.1007/978-3-030-54956-5_1.

[679] F. Å. Nielsen, D. Mietchen, and E. L. Willighagen, "Scholia, scientometrics and wikidata", in *Scientometrics Workshop*, 2017. DOI: 10.1007/978-3-319-70407-4_36.

[680] A. Sinha, Z. Shen, Y. Song, H. Ma, D. Eide, B. P. Hsu, and K. Wang, "An overview of microsoft academic service (MAS) and applications", in *The Web Conference (WWW)*, 2015. DOI: 10.1145/2740908.2742839.

[681] A.-W. Harzing and S. Alakangas, "Google scholar, scopus and the web of science: A longitudinal and cross-disciplinary comparison", *Scientometrics*, vol. 106, no. 2, 2016, pp. 787–804.

[682] M. Färber, "The microsoft academic knowledge graph: A linked data source with 8 billion triples of scholarly data", in *International Semantic Web Conference (ISWC)*, 2019. DOI: 10.1007/978-3-030-30796-7_8.

[683] W. Tunstall-Pedoe, "True knowledge: Open-domain question answering using structured knowledge and inference", *AI Magazine*, vol. 31, no. 3, 2010, pp. 80–92. DOI: 10.1609/aimag.v31i3.2298.

[684] H. Xu, W. Wang, X. Mao, X. Jiang, and M. Lan, "Scaling up open tagging from tens to thousands: Comprehension empowered attribute value extraction from product title", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019.

[685] D. Xu, C. Ruan, E. Körpeoglu, S. Kumar, and K. Achan, "Product knowledge graph embedding for e-commerce", in *ACM Conference on Web Search and Data Mining (WSDM)*, 2020. DOI: 10.1145/3336191.3371778.

[686] A. Pfadler, H. Zhao, J. Wang, L. Wang, P. Huang, and D. L. Lee, "Billion-scale recommendation with heterogeneous side information at taobao", in *IEEE International Conference on Data Engineering (ICDE)*, 2020. DOI: 10.1109/ICDE48307.2020.00148.

[687] E. Weisstein, "Computable data, mathematics, and digital libraries in mathematica and wolfram| alpha", in *Intelligent Computer Mathematics*, Springer, 2014, pp. 26–29.

Full text available at: http://dx.doi.org/10.1561/1900000064

[688] R. Fishkin, "Less than half of google searches now result in a click", in *SparkToro Blog, 13 August 2019, https://sparktoro.com/blog/less-than-half-of-google-searches-now-result-in-a-click/*.

[689] L. Dietz, A. Kotov, and E. Meij, "Utilizing knowledge graphs for text-centric information retrieval (tutorial slides available at https://github.com/laura-dietz/tutorial-utilizing-kg)", in *ACM Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 1387–1390, ACM, 2018. DOI: 10.1145/3209978.3210187.

[690] D. Marantz, "Page zero: A deep dive", in *Microsoft Bing Blogs, 20 September 2013, https://blogs.bing.com/search/2013/09/20/page-zero-a-deep-dive/*.

[691] K. Balog, *Entity-Oriented Search*, vol. 39, ser. The Information Retrieval Series. Springer, 2018. DOI: 10.1007/978-3-319-93935-3.

[692] H. Sun, T. Bedrax-Weiss, and W. W. Cohen, "Pullnet: Open domain question answering with iterative retrieval on knowledge bases and text", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, K. Inui, J. Jiang, V. Ng, and X. Wan, Eds., pp. 2380–2390, Association for Computational Linguistics, 2019. DOI: 10.18653/v1/D19-1242.

[693] D. Chen, A. Fisch, J. Weston, and A. Bordes, "Reading wikipedia to answer open-domain questions", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, R. Barzilay and M. Kan, Eds., pp. 1870–1879, Association for Computational Linguistics, 2017. DOI: 10.18653/v1/P17-1171.

[694] Z. Huang, S. Xu, M. Hu, X. Wang, J. Qiu, Y. Fu, Y. Zhao, Y. Peng, and C. Wang, "Recent trends in deep learning based open-domain textual question answering systems", *IEEE Access*, vol. 8, 2020, pp. 94 341–94 356. DOI: 10.1109/ACCESS.2020.2988903.

[695] B. Oguz, X. Chen, V. Karpukhin, S. Peshterliev, D. Okhonko, M. S. Schlichtkrull, S. Gupta, Y. Mehdad, and S. Yih, "Unified open-domain question answering with structured and unstructured knowledge", *CoRR*, vol. abs/2012.14610, 2020. arXiv: 2012.14610. [Online]. Available: https://arxiv.org/abs/2012.14610.

Full text available at: http://dx.doi.org/10.1561/1900000064

[696] J. Shinavier, K. Branson, W. Zhang, S. Dastgheib, Y. Gao, B. G. Arsintescu, F. özcan, and E. Meij, "Panel: Knowledge graph industry applications", in *The Web Conference (WWW)*, 2019. DOI: 10.1145/3308560.3317711.

[697] M. Rotmensch, Y. Halpern, A. Tlimat, S. Horng, and D. Sontag, "Learning a health knowledge graph from electronic medical records", *Scientific reports*, vol. 7, no. 1, 2017, pp. 1–11.

[698] J. Strötgen, T. Tran, A. Friedrich, D. Milchevski, F. Tomazic, A. Marusczyk, H. Adel, D. Stepanova, F. Hildebrand, and E. Kharlamov, "Towards the bosch materials science knowledge base", in *International Semantic Web Conference (ISWC)*, 2019. [Online]. Available: http://ceur-ws.org/Vol-2456/paper89.pdf.

[699] T. Hubauer, S. Lamparter, P. Haase, and D. M. Herzig, "Use cases of the industrial knowledge graph at siemens", in *International Semantic Web Conference (ISWC)*, 2018. [Online]. Available: http://ceur-ws.org/Vol-2180/paper-86.pdf.

[700] Y. Qi and J. Xiao, "Fintech: AI powers financial services to improve people's lives", *Communication of the ACM*, vol. 61, no. 11, 2018, pp. 65–69. DOI: 10.1145/3239550.

[701] F. Mesquita, M. Cannaviccio, J. Schmidek, P. Mirza, and D. Barbosa, "Knowledgenet: A benchmark dataset for knowledge base population", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019.

[702] R. Meusel, P. Petrovski, and C. Bizer, "The webdatacommons microdata, rdfa and microformat dataset series", in *International Semantic Web Conference (ISWC)*, 2014. DOI: 10.1007/978-3-319-11964-9_18.

[703] F. Petroni, T. Rocktäschel, S. Riedel, P. S. H. Lewis, A. Bakhtin, Y. Wu, and A. H. Miller, "Language models as knowledge bases?", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019. DOI: 10.18653/v1/D19-1250.

[704] Z. Jiang, F. F. Xu, J. Araki, and G. Neubig, "How can we know what language models know", *Transactions of the Association for Computational Linguistics (TACL)*, vol. 8, 2020, pp. 423–438. [Online]. Available: https://transacl.org/ojs/index.php/tacl/article/view/1983.

Full text available at: http://dx.doi.org/10.1561/1900000064

[705] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, and X. Huang, "Pre-trained models for natural language processing: A survey", *CoRR*, vol. abs/2003.08271, 2020.

[706] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of varying shades: Analyzing language in fake news and political fact-checking", in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 2931–2937, 2017. DOI: 10.18653/v1/d17-1317.

[707] K. Popat, S. Mukherjee, J. Strötgen, and G. Weikum, "Where the truth lies: Explaining the credibility of emerging claims on the web and social media", in *The Web Conference (WWW)*, pp. 1003–1012, 2017. DOI: 10.1145/3041021.3055133.

[708] J. Thorne, A. Vlachos, C. Christodoulopoulos, and A. Mittal, "FEVER: a large-scale dataset for fact extraction and verification", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2018. DOI: 10.18653/v1/n18-1074.

[709] Y. Zhang, Z. G. Ives, and D. Roth, "Evidence-based trustworthiness", in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019. DOI: 10.18653/v1/p19-1040.

[710] B. Adair, C. Li, J. Yang, and C. Yu, "Automated pop-up fact-checking: Challenges & progress", in *Computation + Journalism Symposium*, 2019.

[711] G. D. S. Martino, S. Cresci, A. Barron-Cedeño, S. Yu, R. D. Pietro, and P. Nakov, "A survey on computational propaganda detection", in *Joint Conference on Artificial Intelligence (IJCAI)*, pp. 4826–4832, 2020. DOI: 10.24963/ijcai.2020/672.

[712] S. Jiang, S. Baumgartner, A. Ittycheriah, and C. Yu, "Factoring fact-checks: Structured information extraction from fact-checking articles", in *The Web Conference (WWW)*, pp. 1592–1603, 2020. DOI: 10.1145/3366423.3380231.

[713] N. Vedula and S. Parthasarathy, "FACE-KEG: fact checking explained using knowledge graphs", in *ACM Conference on Web Search and Data Mining (WSDM)*, pp. 526–534, ACM, 2021. DOI: 10.1145/3437963.3441828.

Full text available at: http://dx.doi.org/10.1561/1900000064

[714] S. Mukherjee, G. Weikum, and C. Danescu-Niculescu-Mizil, "People on drugs: Credibility of user statements in health communities", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2014. DOI: 10.1145/2623330.2623714.

[715] P. Smeros, C. Castillo, and K. Aberer, "Scilens: Evaluating the quality of scientific news articles using social media and scientific literature indicators", in *The Web Conference (WWW)*, 2019. DOI: 10.1145/3308558.3313657.

[716] D. Wadden, K. Lo, L. L. Wang, S. Lin, M. van Zuylen, A. Cohan, and H. Hajishirzi, "Fact or fiction: Verifying scientific claims", *arXiv:2004.14974*, 2020.

[717] E. Davis, *Representations of commonsense knowledge*. Morgan Kaufmann, 2014.

[718] R. Zellers, Y. Bisk, A. Farhadi, and Y. Choi, "From recognition to cognition: Visual commonsense reasoning", in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. DOI: 10.1109/CVPR.2019.00688.

[719] N. Tandon, G. de Melo, F. Suchanek, and G. Weikum, "Webchild : Harvesting and organizing commonsense knowledge from the web", *ACM Conference on Web Search and Data Mining (WSDM)*, 2014. DOI: 10.1145/2556195.2556245.

[720] N. Tandon, G. de Melo, A. De, and G. Weikum, "Knowlywood: Mining activity knowledge from hollywood narratives", in *ACM Conference on Information and Knowledge Management (CIKM)*, pp. 223–232, 2015. DOI: 10.1145/2806416.2806583.

[721] S. Zhang, R. Rudinger, K. Duh, and B. V. Durme, "Ordinal common-sense inference", *Transactions of the Association for Computational Linguistics (TACL)*, vol. 5, 2017, pp. 379–395. [Online]. Available: https://transacl.org/ojs/index.php/tacl/article/view/1082.

[722] B. D. Mishra, N. Tandon, and P. Clark, "Domain-targeted, high precision knowledge extraction", *Transactions of the Association for Computational Linguistics (TACL)*, vol. 5, 2017, pp. 233–246. [Online]. Available: https://transacl.org/ojs/index.php/tacl/article/view/1064.

Full text available at: http://dx.doi.org/10.1561/1900000064

[723] M. Sap, R. L. Bras, E. Allaway, C. Bhagavatula, N. Lourie, H. Rashkin, B. Roof, N. A. Smith, and Y. Choi, "ATOMIC: an atlas of machine commonsense for if-then reasoning", in *Conference on Artificial Intelligence (AAAI)*, 2019. DOI: 10.1609/aaai.v33i01.33013027.

[724] J. Romero, S. Razniewski, K. Pal, J. Z. Pan, A. Sakhadeo, and G. Weikum, "Commonsense properties from query logs and question answering forums", in *ACM Conference on Information and Knowledge Management (CIKM)*, pp. 1411–1420, 2019. DOI: 10.1145/3357384.3357955.

[725] Y. Chalier, S. Razniewski, and G. Weikum, "Joint reasoning for multi-faceted commonsense knowledge", *International Conference on Automatic Knowledge Base Construction (AKBC)*, 2020.

[726] T.-P. Nguyen, S. Razniewski, and G. Weikum, "Advanced semantics for commonsense knowledge extraction", in *The Web Conference (WWW)*, 2021.

[727] F. Ilievski, A. Oltramari, K. Ma, B. Zhang, D. L. McGuinness, and P. Szekely, "Dimensions of commonsense knowledge", *arXiv preprint arXiv:2101.04640*, 2021.

[728] L. Schubert, "Can we derive general world knowledge from texts", in *International Conference on Human Language Technology (HLT)*, 2002.

[729] J. Gordon and B. Van Durme, "Reporting bias and knowledge acquisition", in *International Workshop on Automated Knowledge Base Construction (AKBC)*, 2013.

[730] P. Singh, T. Lin, E. T. Mueller, G. Lim, T. Perkins, and W. L. Zhu, "Open mind common sense: Knowledge acquisition from the general public", in *OTM Confederated International Conferences*, 2002.

[731] R. Speer, J. Chin, and C. Havasi, "Conceptnet 5.5: An open multilingual graph of general knowledge", in *Conference on Artificial Intelligence (AAAI)*, 2017. [Online]. Available: http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14972.

Full text available at: http://dx.doi.org/10.1561/1900000064

[732] C. Matuszek, M. J. Witbrock, R. C. Kahlert, J. Cabral, D. Schneider, P. Shah, and D. B. Lenat, "Searching for common sense: Populating cyc from the web", in *Conference on Artificial Intelligence (AAAI)*, 2005. [Online]. Available: http://www.aaai.org/Library/AAAI/2005/aaai05-227.php.

[733] M. Sap, V. Shwartz, A. Bosselut, Y. Choi, and D. Roth, "Commonsense reasoning for natural language processing, tutorial with materials at https://homes.cs.washington.edu/~msap/acl2020-commonsense/", A. Savary and Y. Zhang, Eds.

[734] S. Razniewski, N. Tandon, and A. S. Varde, "Information to wisdom: Commonsense knowledge extraction and compilation, tutorial with materials at https://www.mpi-inf.mpg.de/commonsense-tutorial-wsdm-2021", in *ACM Conference on Web Search and Data Mining (WSDM)*, L. Lewin-Eytan, D. Carmel, E. Yom-Tov, E. Agichtein, and E. Gabrilovich, Eds., pp. 1143–1146, ACM, 2021. DOI: 10.1145/3437963.3441664.

[735] F. Ilievski, A. Bosselut, S. Razniewski, and M. Kejriwal, "Commonsense knowledge acquisition and representation, tutorial with materials at https://usc-isi-i2.github.io/AAAI21Tutorial/".

[736] S. T. Dumais, E. Cutrell, J. J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins, "Stuff i've seen: A system for personal information retrieval and re-use", in *ACM Conference on Research and Development in Information Retrieval (SIGIR)*, 2003. DOI: 10.1145/860435.860451.

[737] X. L. Dong and A. Y. Halevy, "A platform for personal information management and integration", in *Conference on Innovative Data Systems Research (CIDR)*, 2005. [Online]. Available: http://cidrdb.org/cidr2005/papers/P10.pdf.

[738] J. Dittrich, M. A. V. Salles, D. Kossmann, and L. Blunschi, "Imemex: Escapes from the personal information jungle", in *Conference on Very Large Databases (VLDB)*, 2005. [Online]. Available: http://www.vldb.org/conf/2005/papers/p1306-dittrich.pdf.

Full text available at: http://dx.doi.org/10.1561/1900000064

[739] D. Montoya, T. P. Tanon, S. Abiteboul, P. Senellart, and F. M. Suchanek, "A knowledge base for personal information management", in *Workshop on Linked Data on the Web (LDOW), co-located with the Web Conference*, 2018. [Online]. Available: http://ceur-ws.org/Vol-2073/article-02.pdf.

[740] K. Balog and T. Kenter, "Personal knowledge graphs: A research agenda", in *ACM Conference on Research and Development in Information Retrieval (SIGIR)*, 2019. DOI: 10.1145/3341981.3344241.

[741] Y. Gurevich and J. M. Wing, "Inverse privacy", *Communication of the ACM*, vol. 59, no. 7, 2016, pp. 38–42. DOI: 10.1145/2838730.

[742] V. Kalokyri, A. Borgida, and A. Marian, "Yourdigitalself: A personal digital trace integration tool", in *ACM Conference on Information and Knowledge Management (CIKM)*, 2018. DOI: 10.1145/3269206.3269219.

[743] R. Awadallah, M. Ramanath, and G. Weikum, "Harmony and dissonance: Organizing the people's voices on political controversies", in *ACM Conference on Web Search and Data Mining (WSDM)*, 2012. DOI: 10.1145/2124295.2124359.

[744] C. Stab and I. Gurevych, "Parsing argumentation structures in persuasive essays", *Computational Linguistics*, vol. 43, no. 3, 2017, pp. 619–659. DOI: 10.1162/COLI_a_00295.

[745] A. Y. Halevy, "The ubiquity of subjectivity", *IEEE Data Engineering Bulletin*, vol. 42, no. 1, 2019, pp. 6–9. [Online]. Available: http://sites.computer.org/debull/A19mar/p6.pdf.

[746] S. Chen, D. Khashabi, W. Yin, C. Callison-Burch, and D. Roth, "Seeing things from a different angle: Discovering diverse perspectives about claims", in *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2019. DOI: 10.18653/v1/n19-1053.

[747] A. Olteanu, C. Castillo, F. Diaz, and E. Kiciman, "Social data: Biases, methodological pitfalls, and ethical boundaries", *Frontiers Big Data*, vol. 2, 2019, p. 13. DOI: 10.3389/fdata.2019.00013.

Full text available at: http://dx.doi.org/10.1561/1900000064

[748] F. M. Suchanek, "The need to move beyond triples", in *Workshop on Narrative Extraction From Texts (Text2Story), co-located with ECIR Conference*, 2020. [Online]. Available: http://ceur-ws.org/Vol-2593/paper12.pdf.

[749] N. Bhutani, A. Traylor, C. Chen, X. Wang, B. Golshan, and W.-C. Tan, "Sampo: Unsupervised knowledge base construction for opinions and implications", in *Automatic Knowledge Base Construction (AKBC)*, 2020.

[750] K. A. Khatib, Y. Hou, H. Wachsmuth, C. Jochim, F. Bonin, and B. Stein, "End-to-end argumentation knowledge graph construction", in *Conference on Artificial Intelligence (AAAI)*, 2020. [Online]. Available: https://aaai.org/ojs/index.php/AAAI/article/view/6231.

[751] S. Sarawagi and S. Chakrabarti, "Open-domain quantity queries on web tables: Annotation, response, and consensus models", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 711–720, 2014. DOI: 10.1145/2623330.2623749.

[752] V. T. Ho, Y. Ibrahim, K. Pal, K. Berberich, and G. Weikum, "Qsearch: Answering quantity queries from text", in *International Semantic Web Conference (ISWC)*, 2019. DOI: 10.1007/978-3-030-30793-6_14.

[753] G. Weikum, "Entities with quantities", *IEEE Data Engineering Bulletin*, vol. 43, no. 1, 2020, pp. 4–8. [Online]. Available: http://sites.computer.org/debull/A20mar/p4.pdf.

[754] S. Razniewski and P. Das, "Structured knowledge: Have we made progress? an extrinsic study of KB coverage over 19 years", in *ACM Conference on Information and Knowledge Management (CIKM)*, pp. 3317–3320, ACM, 2020. DOI: 10.1145/3340531.3417447.

[755] R. Speer, J. Chin, and C. Havasi, "Conceptnet 5.5: An open multilingual graph of general knowledge", in *Conference on Artificial Intelligence (AAAI)*, 2017. [Online]. Available: http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14972.

[756] L. von Ahn and L. Dabbish, "Labeling images with a computer game", in *Conference on Human Factors in Computing Systems CHI*, pp. 319–326, ACM, 2004. DOI: 10.1145/985692.985733.

Full text available at: http://dx.doi.org/10.1561/1900000064

[757] L. Von Ahn, M. Kedia, and M. Blum, "Verbosity: A game for collecting common-sense facts", in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 75–78, 2006.