

References

1. Achichi, M., Cheatham, M., Dragisic, Z., Euzenat, J., Faria, D., Ferrara, A., Flouris, G., Fundulaki, I., Harrow, I., Ivanova, V., et al.: Results of the ontology alignment evaluation initiative 2016. In: OM: Ontology Matching, pp. 73–129. No commercial editor. (2016)
2. Agichtein, E., Gravano, L.: Snowball: extracting relations from large plain-text collections. In: Proceedings of the Fifth ACM Conference on Digital Libraries, pp. 85–94. ACM (2000)
3. Ahn, D.: The stages of event extraction. In: Proceedings of the Workshop on Annotating and Reasoning About Time and Events, pp. 1–8. Association for Computational Linguistics (2006)
4. Alfonseca, E., Manandhar, S.: An unsupervised method for general named entity recognition and automated concept discovery. In: Proceedings of the 1st International Conference on General WordNet, Mysore, pp. 34–43 (2002)
5. Allemang, D., Hendler, J.: Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL. Elsevier, Amsterdam (2011)
6. Arasu, A., Garcia-Molina, H.: Extracting structured data from web pages. In: Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, pp. 337–348. ACM (2003)
7. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.: Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**(1), 25 (2000)
8. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: Dbpedia: a nucleus for a web of open data. In: The Semantic Web, pp. 722–735. Springer, Berlin (2007)
9. Bach, N., Badaskar, S.: A survey on relation extraction. Language Technologies Institute, Carnegie Mellon University (2007)
10. Banko, M., Cafarella, M.J., Soderland, S., Broadhead, M., Etzioni, O.: Open information extraction from the web. In: IJCAI, vol. 7, pp. 2670–2676 (2007)
11. Bauer, F., Kaltenböck, M.: Linked Open Data: The Essentials. Edition mono/monochrom, Vienna (2011)
12. Baxter, R., Christen, P., Churches, T., et al.: A comparison of fast blocking methods for record linkage. In: ACM SIGKDD, vol. 3, pp. 25–27. Citeseer (2003)
13. Benajiba, Y., Diab, M., Rosso, P.: Arabic named entity recognition using optimized feature sets. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 284–293. Association for Computational Linguistics (2008)
14. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(8), 1798–1828 (2013)

15. Benjelloun, O., Garcia-Molina, H., Menestrina, D., Su, Q., Whang, S.E., Widom, J.: Swoosh: a generic approach to entity resolution. *VLDB J.: Int. J. Very Large Data Bases* **18**(1), 255–276 (2009)
16. Berant, J., Srikumar, V., Chen, P.C., Vander Linden, A., Harding, B., Huang, B., Clark, P., Manning, C.D.: Modeling biological processes for reading comprehension. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1499–1510 (2014)
17. Berners-Lee, T., Fielding, R., Masinter, L.: Uniform resource identifier (URI): generic syntax. Technical report (2004)
18. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. *Sci. Am.* **284**(5), 34–43 (2001)
19. Bhaskaran, S., Rafeeq, P.: A survey on relation extraction methodologies from unstructured text. In: *Emerging Trends in Engineering, Science and Technology for Society, Energy and Environment*, pp. 869–874. CRC Press, Leiden (2018)
20. Bhattacharya, I., Getoor, L.: A latent dirichlet model for unsupervised entity resolution. In: *Proceedings of the 2006 SIAM International Conference on Data Mining*, pp. 47–58. SIAM (2006)
21. Bhattacharya, I., Getoor, L.: Collective entity resolution in relational data. *ACM Trans. Knowl. Discov. Data (TKDD)* **1**(1), 5 (2007)
22. Bick, E.: A named entity recognizer for Danish. In: *LREC. Citeseer* (2004)
23. Bilenko, M., Mooney, R.J.: Adaptive duplicate detection using learnable string similarity measures. In: *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 39–48. ACM (2003)
24. Bilke, A., Naumann, F.: Schema matching using duplicates. In: *21st International Conference on Data Engineering, 2005 (ICDE 2005). Proceedings*, pp. 69–80. IEEE (2005)
25. Bizer, C.: The emerging web of linked data. *IEEE Intell. Syst.* **24**(5), 87–92 (2009)
26. Bizer, C., Heath, T., Berners-Lee, T.: Linked data: the story so far. In: *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, pp. 205–227. IGI Global (2011)
27. Bizer, C., Heath, T., Idehen, K., Berners-Lee, T.: Linked data on the web (ldow2008). In: *Proceedings of the 17th International Conference on World Wide Web*, pp. 1265–1266. ACM (2008)
28. Björne, J., Heimonen, J., Ginter, F., Airola, A., Pahikkala, T., Salakoski, T.: Extracting complex biological events with rich graph-based feature sets. In: *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task*, pp. 10–18. Association for Computational Linguistics (2009)
29. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003)
30. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: *Advances in Neural Information Processing Systems*, pp. 2787–2795 (2013)
31. Bordes, A., Weston, J., Collobert, R., Bengio, Y., et al.: Learning structured embeddings of knowledge bases. In: *AAAI*, vol. 6, p. 6 (2011)
32. Brin, S.: Extracting patterns and relations from the world wide web. In: *International Workshop on the World Wide Web and Databases*, pp. 172–183. Springer (1998)
33. Cao, Y., Chen, Z., Zhu, J., Yue, P., Lin, C.Y., Yu, Y.: Leveraging unlabeled data to scale blocking for record linkage. In: *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, vol. 22, p. 2211 (2011)
34. Chakrabarti, K., Chaudhuri, S., Cheng, T., Xin, D.: A framework for robust discovery of entity synonyms. In: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1384–1392. ACM (2012)
35. Chambers, N., Jurafsky, D.: Unsupervised learning of narrative event chains. In: *Proceedings of ACL-08: HLT*, pp. 789–797 (2008)
36. Chang, C.H., Kayed, M., Girgis, M.R., Shaalan, K.F.: A survey of web information extraction systems. *IEEE Trans. Knowl. Data Eng.* **18**(10), 1411–1428 (2006)

37. Chang, C.H., Kuo, S.C.: Olera: semisupervised web-data extraction with visual support. *IEEE Intell. Syst.* **19**(6), 56–64 (2004)
38. Chang, C.H., Lui, S.C.: Iepad: information extraction based on pattern discovery. In: *Proceedings of the 10th International Conference on World Wide Web*, pp. 681–688. ACM (2001)
39. Christen, P.: Automatic record linkage using seeded nearest neighbour and support vector machine classification. In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 151–159. ACM (2008)
40. Christen, P.: Febrl-: an open source data cleaning, deduplication and record linkage system with a graphical user interface. In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1065–1068. ACM (2008)
41. Christen, P.: *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Springer, Heidelberg (2012)
42. Christen, P.: A survey of indexing techniques for scalable record linkage and deduplication. *IEEE Trans. Knowl. Data Eng.* **24**(9), 1537–1555 (2012)
43. Coates-Stephens, S.: The analysis and acquisition of proper names for the understanding of free text. *Comput. Humanit.* **26**(5–6), 441–456 (1992)
44. Cohen, W.W., Sarawagi, S.: Exploiting dictionaries in named entity extraction: combining semi-Markov extraction processes and data integration methods. In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 89–98. ACM (2004)
45. Collins, M.: Ranking algorithms for named-entity extraction: boosting and the voted perceptron. In: *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 489–496. Association for Computational Linguistics (2002)
46. Collins, M., Singer, Y.: Unsupervised models for named entity classification. In: *1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora* (1999)
47. Collobert, R., Weston, J.: A unified architecture for natural language processing: deep neural networks with multitask learning. In: *Proceedings of the 25th International Conference on Machine Learning*, pp. 160–167. ACM (2008)
48. Cowie, J., Lehnert, W.: Information extraction. *Commun. ACM* **39**(1), 80–91 (1996)
49. Crescenzi, V., Mecca, G., Merialdo, P., et al.: Roadrunner: towards automatic data extraction from large web sites. In: *VLDB*, vol. 1, pp. 109–118 (2001)
50. Cunningham, H.: Information extraction, automatic. In: Brown, K. (eds.) *Encyclopedia of Language & Linguistics*, vol. 5, 2nd edn., pp. 665–677. Elsevier, Oxford (2006)
51. Dai, A.M., Olah, C., Le, Q.V.: Document embedding with paragraph vectors. *arXiv preprint arXiv:1507.07998* (2015)
52. Datar, M., Immorlica, N., Indyk, P., Mirrokni, V.S.: Locality-sensitive hashing scheme based on p-stable distributions. In: *Proceedings of the Twentieth Annual Symposium on Computational Geometry*, pp. 253–262. ACM (2004)
53. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. *J. R. Stat. Soc. Ser. B (Methodol.)* **39**(1), 1–22 (1977)
54. Do, Q.X., Lu, W., Roth, D.: Joint inference for event timeline construction. In: *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 677–687. Association for Computational Linguistics (2012)
55. Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmman, T., Sun, S., Zhang, W.: Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 601–610. ACM (2014)
56. Duan, S., Fokoue, A., Hassanzadeh, O., Kementsietsidis, A., Srinivas, K., Ward, M.J.: Instance-based matching of large ontologies using locality-sensitive hashing. In: *International Semantic Web Conference*, pp. 49–64. Springer (2012)

57. Elmagarmid, A.K., Ipeirotis, P.G., Verykios, V.S.: Duplicate record detection: a survey. *IEEE Trans. Knowl. Data Eng.* **19**(1), 1–16 (2007)
58. Esteban, C., Tresp, V., Yang, Y., Baier, S., Krompaß, D.: Predicting the co-evolution of event and knowledge graphs. In: 2016 19th International Conference on Information Fusion (FUSION), pp. 98–105. Ieee (2016)
59. Etzioni, O., Cafarella, M., Downey, D., Popescu, A.M., Shaked, T., Soderland, S., Weld, D.S., Yates, A.: Unsupervised named-entity extraction from the web: an experimental study. *Artif. Intell.* **165**(1), 91–134 (2005)
60. Euzenat, J., Shvaiko, P., et al.: *Ontology Matching*, vol. 18. Springer, Berlin (2007)
61. Fellegi, I.P., Sunter, A.B.: A theory for record linkage. *J. Am. Stat. Assoc.* **64**(328), 1183–1210 (1969)
62. Feng, J., Huang, M., Yang, Y., et al.: Gake: graph aware knowledge embedding. In: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pp. 641–651 (2016)
63. Ferraram, A., Nikolov, A., Scharffe, F.: Data linking for the semantic web. *Semant. Web: Ontol. Knowl. Base Enabled Tools Serv. Appl.* **169**, 326 (2013)
64. Gaizauskas, R., Wakao, T., Humphreys, K., Cunningham, H., Wilks, Y.: University of sheffield: description of the lasie system as used for MUC-6. Technical report, Sheffield University (UK) (1995)
65. Gao, J., Galley, M., Li, L.: Neural approaches to conversational AI. In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 1371–1374. ACM (2018)
66. Getoor, L., Machanavajjhala, A.: Entity resolution: theory, practice & open challenges. *Proc. VLDB Endowment* **5**(12), 2018–2019 (2012)
67. Graham, W. (2012). Facebook developer tools. In: *Beginning Facebook Game Apps Development*, pp. 201–229. Apress
68. Grishman, R., Sundheim, B.: Message understanding conference-6: a brief history. In: *COLING 1996: The 16th International Conference on Computational Linguistics*, vol. 1 (1996)
69. Guha, R.V., Brickley, D., Macbeth, S.: Schema.org: evolution of structured data on the web. *Commun. ACM* **59**(2), 44–51 (2016)
70. Guo, S., Wang, Q., Wang, B., Wang, L., Guo, L.: Semantically smooth knowledge graph embedding. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, vol. 1, pp. 84–94 (2015)
71. Guo, S., Wang, Q., Wang, L., Wang, B., Guo, L.: Jointly embedding knowledge graphs and logical rules. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 192–202 (2016)
72. Gusfield, D., Irving, R.W.: *The Stable Marriage Problem: Structure and Algorithms*. MIT Press, Cambridge (1989)
73. Hachey, B., Radford, W., Nothman, J., Honnibal, M., Curran, J.R.: Evaluating entity linking with wikipedia. *Artif. Intell.* **194**, 130–150 (2013)
74. Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora. In: *Proceedings of the 14th Conference on Computational Linguistics*, vol. 2, pp. 539–545. Association for Computational Linguistics (1992)
75. Heath, T., Bizer, C.: Linked data: evolving the web into a global data space. *Synth. Lect. Semant. Web: Theory Technol.* **1**(1), 1–136 (2011)
76. Hernández, M.A., Stolfo, S.J.: Real-world data is dirty: data cleansing and the merge/purge problem. *Data Min. Knowl. Discov.* **2**(1), 9–37 (1998)
77. Hogue, A., Karger, D.: Thresher: automating the unwrapping of semantic content from the world wide web. In: *Proceedings of the 14th International Conference on World Wide Web*, pp. 86–95. ACM (2005)

78. Isozaki, H., Kazawa, H.: Efficient support vector classifiers for named entity recognition. In: Proceedings of the 19th International Conference on Computational Linguistics, vol. 1, pp. 1–7. Association for Computational Linguistics (2002)
79. Jain, A.K., Dubes, R.C.: Algorithms for Clustering Data. Prentice-Hall, Englewood Cliffs (1988)
80. Jansche, M.: Named entity extraction with conditional Markov models and classifiers. In: Proceedings of the 6th Conference on Natural Language Learning, vol. 20, pp. 1–4. Association for Computational Linguistics (2002)
81. Ji, H., Grishman, R.: Refining event extraction through cross-document inference. Proc. ACL-08: HLT, pp. 254–262 (2008)
82. Jiang, T., Liu, T., Ge, T., Sha, L., Li, S., Chang, B., Sui, Z.: Encoding temporal information for time-aware link prediction. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 2350–2354 (2016)
83. Jiang, X., Tresp, V., Huang, Y., Nickel, M.: Link prediction in multi-relational graphs using additive models. *SeRSy* **919**, 1–12 (2012)
84. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759 (2016)
85. Kejriwal, M.: Populating entity name systems for big data integration. In: International Semantic Web Conference, pp. 521–528. Springer (2014)
86. Kejriwal, M.: Populating a Linked Data Entity Name System: A Big Data Solution to Unsupervised Instance Matching, vol. 27. IOS Press, Amsterdam (2016)
87. Kejriwal, M., Miranker, D.P.: An unsupervised algorithm for learning blocking schemes. In: 2013 IEEE 13th International Conference on Data Mining (ICDM), pp. 340–349. IEEE (2013)
88. Kejriwal, M., Miranker, D.P.: A two-step blocking scheme learner for scalable link discovery. In: Proceedings of the 9th International Conference on Ontology Matching, vol. 1317, pp. 49–60. CEUR-WS.org (2014)
89. Kejriwal, M., Miranker, D.P.: An unsupervised instance matcher for schema-free rdf data. *Web Semant.: Sci. Serv. Agents World Wide Web* **35**, 102–123 (2015)
90. Kejriwal, M., Szekely, P.: Neural embeddings for populated geonames locations. In: International Semantic Web Conference, pp. 139–146. Springer (2017)
91. Kim, J.D., Ohta, T., Pyysalo, S., Kano, Y., Tsujii, J.: Overview of bionlp’09 shared task on event extraction. In: Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task, pp. 1–9. Association for Computational Linguistics (2009)
92. Kimmig, A., Bach, S., Broecheler, M., Huang, B., Getoor, L.: A short introduction to probabilistic soft logic. In: Proceedings of the NIPS Workshop on Probabilistic Programming: Foundations and Applications, pp. 1–4 (2012)
93. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms, vol. 8. Springer, Cham (2013)
94. Kolb, L., Thor, A., Rahm, E.: Dedoop: efficient deduplication with hadoop. Proc. VLDB Endowment **5**(12), 1878–1881 (2012)
95. Kolb, L., Thor, A., Rahm, E.: Multi-pass sorted neighborhood blocking with mapreduce. *Comput. Sci.-Res. Develop.* **27**(1), 45–63 (2012)
96. Köpcke, H., Rahm, E.: Frameworks for entity matching: a comparison. *Data Knowl. Eng.* **69**(2), 197–210 (2010)
97. Kudo, T., Matsumoto, Y.: Chunking with support vector machines. In: Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies, pp. 1–8. Association for Computational Linguistics (2001)
98. Kumar, S.: A survey of deep learning methods for relation extraction. arXiv preprint arXiv:1705.03645 (2017)
99. Kushmerick, N., Weld, D.S., Doorenbos, R.: Wrapper Induction for Information Extraction, pp. 729–737. University of Washington, Washington (1997)

100. Lafferty, J., McCallum, A., Pereira, F.C.: Proceedings of the Eighteenth International Conference on Machine Learning (ICML '01), pp. 282–289. Morgan Kaufmann Publishers Inc., San Francisco (2001)
101. Leetaru, K., Schrodtt, P.A.: Gdelt: global data on events, location, and tone, 1979–2012. In: ISA Annual Convention, vol. 2, pp. 1–49. Citeseer (2013)
102. Li, Q., Ji, H., Huang, L.: Joint event extraction via structured prediction with global features. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, pp. 73–82 (2013)
103. Li, X., Morie, P., Roth, D.: Identification and tracing of ambiguous names: discriminative and generative approaches. In: Proceedings of the National Conference on Artificial Intelligence, 1999, pp. 419–424. AAAI Press/MIT Press, Menlo Park/Cambridge/London (2004)
104. Liao, S., Grishman, R.: Using document level cross-event inference to improve event extraction. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 789–797. Association for Computational Linguistics (2010)
105. Lin, Y., Liu, Z., Sun, M.: Knowledge representation learning with entities, attributes and relations. *Ethnicity* **1**, 41–52 (2016)
106. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: AAAI, vol. 15, pp. 2181–2187 (2015)
107. Liu, X., Zhang, S., Wei, F., Zhou, M.: Recognizing named entities in tweets. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, vol. 1, pp. 359–367. Association for Computational Linguistics (2011)
108. Maaten, L.v.d., Hinton, G.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008)
109. Manning, C.D., Manning, C.D., Schütze, H.: Foundations of Statistical Natural Language Processing. MIT Press, Cambridge (1999)
110. Martínez, V., Berzal, F., Cubero, J.C.: A survey of link prediction in complex networks. *ACM Comput. Surv. (CSUR)* **49**(4), 69 (2017)
111. McCallum, A.: Information extraction: distilling structured data from unstructured text. *Queue* **3**(9), 4 (2005)
112. McCallum, A., Nigam, K., Ungar, L.H.: Efficient clustering of high-dimensional data sets with application to reference matching. In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 169–178. ACM (2000)
113. McClosky, D., Manning, C.D.: Learning constraints for consistent timeline extraction. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 873–882. Association for Computational Linguistics (2012)
114. McClosky, D., Surdeanu, M., Manning, C.D.: Event extraction as dependency parsing. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, vol. 1, pp. 1626–1635. Association for Computational Linguistics (2011)
115. McDonald, D.: Internal and external evidence in the identification and semantic categorization of proper names. Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics, 22–26 June 1993, Ohio State University, Columbus. ACL (1993)
116. Mena, E., Kashyap, V., Illarramendi, A., Sheth, A.: Domain specific ontologies for semantic information brokering on the global information infrastructure. In: Formal Ontology in Information Systems, vol. 46, pp. 269–283. IOS Press, Amsterdam (1998)
117. Mikheev, A., Moens, M., Grover, C.: Named entity recognition without gazetteers. In: Proceedings of the Ninth Conference on European Chapter of the Association for Computational Linguistics, pp. 1–8. Association for Computational Linguistics (1999)
118. Mikolov, T., Chen, K., Corrado, G., Dean, J., Sutskever, L., Zweig, G.: word2vec. <https://code.google.com/p/word2vec> (2013)

119. Mooney, R.J., Bunescu, R.: Mining knowledge from text using information extraction. *ACM SIGKDD Explor. Newslett.* **7**(1), 3–10 (2005)
120. Nadeau, D.: Semi-supervised named entity recognition: learning to recognize 100 entity types with little supervision. Ph.D. thesis, University of Ottawa (2007)
121. Nadeau, D., Sekine, S.: A survey of named entity recognition and classification. *Linguisticae Investigationes* **30**(1), 3–26 (2007)
122. Ng, V.: Unsupervised models for conference resolution. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 640–649 (2008). Association for Computational Linguistics
123. Newcombe, H.B., Kennedy, J.M., Axford, S., James, A.P.: Automatic linkage of vital records. *Science* **130**(3381), 954–959 (1959)
124. Ngomo, A.C.N., Auer, S.: Limes-a time-efficient approach for large-scale link discovery on the web of data. In: *IJCAI*, pp. 2312–2317 (2011)
125. Ngomo, A.C.N., Lehmann, J., Auer, S., Höffner, K.: Raven-active learning of link specifications. In: *Proceedings of the 6th International Conference on Ontology Matching*, vol. 814, pp. 25–36. CEUR-WS.org (2011)
126. Ngomo, A.C.N., Lyko, K.: Eagle: efficient active learning of link specifications using genetic programming. In: *Extended Semantic Web Conference*, pp. 149–163. Springer (2012)
127. Ngomo, A.C.N., Lyko, K.: Unsupervised learning of link specifications: deterministic vs. non-deterministic. In: *Proceedings of the 8th International Conference on Ontology Matching*, vol. 1111, pp. 25–36. CEUR-WS.org (2013)
128. Ngomo, A.C.N., Lyko, K., Christen, V.: Coala-correlation-aware active learning of link specifications. In: *Extended Semantic Web Conference*, pp. 442–456. Springer (2013)
129. Nguyen, D.Q.: An overview of embedding models of entities and relationships for knowledge base completion. *arXiv preprint arXiv:1703.08098* (2017)
130. Nguyen, T.H., Grishman, R.: Relation extraction: perspective from convolutional neural networks. In: *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, pp. 39–48 (2015)
131. Nickel, M., Tresp, V., Krieger, H.P.: Factorizing YAGO: scalable machine learning for linked data. In: *Proceedings of the 21st International Conference on World Wide Web*, pp. 271–280. ACM (2012)
132. Nie, F., Zhu, W., Li, X.: Unsupervised large graph embedding. In: *AAAI*, pp. 2422–2428 (2017)
133. Novak, B.: A survey of focused web crawling algorithms. In: *Proceedings of SIKDD*, pp. 55–58. Citeseer (2004)
134. Palmer, D.D., Day, D.S.: A statistical profile of the named entity task. In: *Proceedings of the Fifth Conference on Applied Natural Language Processing*, pp. 190–193. Association for Computational Linguistics (1997)
135. Pan, J.Z.: Resource description framework. In: *Handbook on Ontologies*, pp. 71–90. Springer, Berlin/Heidelberg (2009)
136. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**(10), 1345–1359 (2010)
137. Papadakis, G., Ioannou, E., Palpanas, T., Niederee, C., Nejdl, W.: A blocking framework for entity resolution in highly heterogeneous information spaces. *IEEE Trans. Knowl. Data Eng.* **25**(12), 2665–2682 (2013)
138. Patrick, J., Whitelaw, C., Munro, R.: Slinerc: the sydney language-independent named entity recogniser and classifier. In: *Proceedings of the 6th Conference on Natural Language Learning*, vol. 20, pp. 1–4. Association for Computational Linguistics (2002)
139. Paulheim, H.: Knowledge graph refinement: a survey of approaches and evaluation methods. *Semant. Web* **8**(3), 489–508 (2017)
140. Pennington, J., Socher, R., Manning, C.: Glove: global vectors for word representation. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543 (2014)

141. Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: online learning of social representations. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 701–710. ACM (2014)
142. Piskorski, J., Tanev, H., Atkinson, M., Van Der Goot, E., Zavarella, V.: Online news event extraction for global crisis surveillance. In: Transactions on Computational Collective Intelligence V, pp. 182–212. Springer, Berlin/New York (2011)
143. Poibeau, T.: Dealing with metonymic readings of named entities. arXiv preprint cs/0607052 (2006)
144. Pujara, J., Augustine, E., Getoor, L.: Sparsity and noise: where knowledge graph embeddings fall short. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 1751–1756 (2017)
145. Pujara, J., Miao, H., Getoor, L., Cohen, W.: Knowledge graph identification. In: International Semantic Web Conference, pp. 542–557. Springer (2013)
146. Raghavan, H., Allan, J., McCallum, A.: An exploration of entity models, collective classification and relation description. In: KDD Workshop on Link Analysis and Group Detection, pp. 1–10 (2004)
147. Rahm, E., Bernstein, P.A.: A survey of approaches to automatic schema matching. *VLDB J.* **10**(4), 334–350 (2001)
148. Ramadan, B., Christen, P.: Unsupervised blocking key selection for real-time entity resolution. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 574–585. Springer (2015)
149. Ratnov, L., Roth, D.: Design challenges and misconceptions in named entity recognition. In: Proceedings of the Thirteenth Conference on Computational Natural Language Learning, pp. 147–155. Association for Computational Linguistics (2009)
150. Ravikumar, P., Cohen, W.W.: A hierarchical graphical model for record linkage. In: Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, pp. 454–461. AUAI Press (2004)
151. Ravin, Y., Leacock, C.: Polysemy: Theoretical and Computational Approaches. OUP, Oxford (2000)
152. Ravin, Y., Wacholder, N.: Extracting Names from Natural-Language Text. Citeseer, University Park (1997)
153. Raytheon, B.: Technologies. BBN accent event coding evaluation. Technical report (2015)
154. Riedel, S., McCallum, A.: Fast and robust joint models for biomedical event extraction. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 1–12. Association for Computational Linguistics (2011)
155. Riloff, E., Jones, R., et al.: Learning dictionaries for information extraction by multi-level bootstrapping. In: AAAI/IAAI, pp. 474–479 (1999)
156. Ritter, A., Clark, S., Etzioni, O., et al.: Named entity recognition in tweets: an experimental study. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 1524–1534. Association for Computational Linguistics (2011)
157. Rocktäschel, T., Singh, S., Riedel, S.: Injecting logical background knowledge into embeddings for relation extraction. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1119–1129 (2015)
158. Rocktäschel, T., Weidlich, M., Leser, U.: Chemspot: a hybrid system for chemical named entity recognition. *Bioinformatics* **28**(12), 1633–1640 (2012)
159. Rong, S., Niu, X., Xiang, E.W., Wang, H., Yang, Q., Yu, Y.: A machine learning approach for instance matching based on similarity metrics. In: International Semantic Web Conference, pp. 460–475. Springer (2012)
160. Salton, G., McGill, M.J.: Introduction to Modern Information Retrieval. McGraw-Hill, New York (1986)
161. Sarawagi, S., et al.: Information extraction. *Found. Trends® Databases* **1**(3), 261–377 (2008)

162. Settles, B.: Biomedical named entity recognition using conditional random fields and rich feature sets. In: *Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and Its Applications*, pp. 104–107. Association for Computational Linguistics (2004)
163. Shadbolt, N., Berners-Lee, T., Hall, W.: The semantic web revisited. *IEEE Intell. Syst.* **21**(3), 96–101 (2006)
164. Singhal, A.: Introducing the knowledge graph: things, not strings. *Off. Google Blog* **5** (2012). <https://www.blog.google/products/search/introducing-knowledge-graph-things-not/>
165. Socher, R., Chen, D., Manning, C.D., Ng, A.: Reasoning with neural tensor networks for knowledge base completion. In: *Advances in Neural Information Processing Systems*, pp. 926–934 (2013)
166. Song, D., Heflin, J.: Automatically generating data linkages using a domain-independent candidate selection approach. In: *The Semantic Web—ISWC 2011*, pp. 649–664 (2011)
167. Suchanek, F.M., Kasneci, G., Weikum, G.: YAGO: a core of semantic knowledge. In: *Proceedings of the 16th International Conference on World Wide Web*, pp. 697–706. ACM (2007)
168. Tay, Y., Luu, A.T., Hui, S.C., Brauer, F.: Random semantic tensor ensemble for scalable knowledge graph link prediction. In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 751–760. ACM (2017)
169. Thielen, C.: An approach to proper name tagging for German. *arXiv preprint cmp-lg/9506024* (1995)
170. Tommasi, T., Patricia, N., Caputo, B., Tuytelaars, T.: A deeper look at dataset bias. In: *Domain Adaptation in Computer Vision Applications*, pp. 37–55. Springer, Cham (2017)
171. Tsuruoka, Y., Tsujii, J.: Boosting precision and recall of dictionary-based protein name recognition. In: *Proceedings of the ACL 2003 Workshop on Natural Language Processing in Biomedicine*, vol. 13, pp. 41–48. Association for Computational Linguistics (2003)
172. Volz, J., Bizer, C., Gaedke, M., Kobilarov, G.: Silk—a link discovery framework for the web of data. *LDOW* **538** (2009). <http://ceur-ws.org/Vol-538/>
173. Voutilainen, A.: Part-of-speech tagging. In: *The Oxford Handbook of Computational Linguistics*, pp. 219–232. Oxford University Press, Oxford (2003)
174. Vrandečić, D., Krötzsch, M.: Wikidata: a free collaborative knowledgebase. *Commun. ACM* **57**(10), 78–85 (2014)
175. Wactlar, H.D.: New directions in video information extraction and summarization. In: *Proceedings of the 10th DELOS Workshop*, Sanorini, pp. 24–25 (1999)
176. Wang, Q., Mao, Z., Wang, B., Guo, L.: Knowledge graph embedding: a survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.* **29**(12), 2724–2743 (2017)
177. Wang, Q., Wang, B., Guo, L., et al.: Knowledge base completion using embeddings and rules. In: *IJCAI*, pp. 1859–1866 (2015)
178. Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph and text jointly embedding. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1591–1601 (2014)
179. Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: *AAAI*, vol. 14, pp. 1112–1119 (2014)
180. Ward, M.D., Beger, A., Cutler, J., Dickenson, M., Dorff, C., Radford, B.: Comparing gdelt and icews event data. *Analysis* **21**(1), 267–97 (2013)
181. Wei, Z., Zhao, J., Liu, K., Qi, Z., Sun, Z., Tian, G.: Large-scale knowledge base completion: inferring via grounding network sampling over selected instances. In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 1331–1340. ACM (2015)
182. Weld, D.S., Bansal, G.: Intelligible artificial intelligence. *arXiv preprint arXiv:1803.04263* (2018)
183. Winkler, W.E.: Improved Decision Rules in the Fellegi-Sunter Model of Record Linkage, pp. 274–279. Bureau of the Census (1993)

184. Winkler, W.E.: Methods for record linkage and Bayesian networks. Technical report, Statistical Research Division, US Census Bureau, Washington, DC (2002)
185. Xie, R., Liu, Z., Jia, J., Luan, H., Sun, M.: Representation learning of knowledge graphs with entity descriptions. In: AAAI, pp. 2659–2665 (2016)
186. Xie, R., Liu, Z., Sun, M.: Representation learning of knowledge graphs with hierarchical types. In: IJCAI, pp. 2965–2971 (2016)
187. Yates, A., Cafarella, M., Banko, M., Etzioni, O., Broadhead, M., Soderland, S.: Textrunner: open information extraction on the web. In: Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, pp. 25–26. Association for Computational Linguistics (2007)
188. Yu, S., Bai, S., Wu, P.: Description of the kent ridge digital labs system used for MUC-7. In: Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference Held in Fairfax (1998)
189. Zelenko, D., Aone, C., Richardella, A.: Kernel methods for relation extraction. *J. Mach. Learn. Res.* **3**, 1083–1106 (2003)
190. Zheng, S., Hao, Y., Lu, D., Bao, H., Xu, J., Hao, H., Xu, B.: Joint entity and relation extraction based on a hybrid neural network. *Neurocomputing* **257**, 59–66 (2017)
191. Zhou, G., Su, J.: Named entity recognition using an HMM-based chunk tagger. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 473–480. Association for Computational Linguistics (2002)