

REFERENCES

- [1] Y. C. Rathod, *Methods and systems for brands social networks (bsn) platform*. Google Patents, 2011.
- [2] V. N. Gudivada, D. Rao, and V. V. Raghavan, “NoSQL systems for big data management,” in *2014 IEEE World congress on services*, 2014, pp. 190–197.
- [3] M. J. Denny and A. Spirling, “Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It,” *Political Analysis*, vol. 26, no. 2, pp. 168–189, Apr. 2018.
- [4] F. Sun, A. Belatreche, S. Coleman, T. M. McGinnity, and Y. Li, “Pre-processing online financial text for sentiment classification: A natural language processing approach,” in *2014 IEEE Conference on Computational Intelligence for Financial Engineering Economics (CIFER)*, 2014, pp. 122–129.
- [5] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, “The Stanford CoreNLP natural language processing toolkit,” in *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, 2014, pp. 55–60.
- [6] D. Nadeau and S. Sekine, “A survey of named entity recognition and classification,” *Linguisticae Investigationes*, vol. 30, no. 1, pp. 3–26, 2007.
- [7] “The Porter stemming algorithm: then and now | Program | Vol 40, No 3.” [Online]. Available: <https://www.emeraldinsight.com/doi/abs/10.1108/00330330610681295>. [Accessed: 23-Jun-2018].
- [8] C. D. Paice, “An Evaluation Method for Stemming Algorithms,” in *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, 1994, pp. 42–50.
- [9] R. R. Larson, “Introduction to Information Retrieval,” *Journal of the American Society for Information Science and Technology*, vol. 61, no. 4, pp. 852–853.
- [10] G. V. Bard, “Spelling-error tolerant, order-independent pass-phrases via the Damerau-Levenshtein string-edit distance metric,” in *Proceedings of the fifth Australasian symposium on ACSW frontiers-Volume 68*, 2007, pp. 117–124.
- [11] E. Brill and R. C. Moore, “An Improved Error Model for Noisy Channel Spelling Correction,” in *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, Stroudsburg, PA, USA, 2000, pp. 286–293.
- [12] “Comparison of biosequences - ScienceDirect.” [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0196885881900464>. [Accessed: 26-Jun-2018].
- [13] M. U. S. Shameem and R. Ferdous, “An efficient k-means algorithm integrated with Jaccard distance measure for document clustering,” in *2009 First Asian Himalayas International Conference on Internet*, 2009, pp. 1–6.
- [14] W. E. Winkler, “Overview of record linkage and current research directions,”

BUREAU OF THE CENSUS, 2006.

- [15] B. Mohit, “Named Entity Recognition,” in *Natural Language Processing of Semitic Languages*, Springer, Berlin, Heidelberg, 2014, pp. 221–245.
- [16] G. Zhou and J. Su, “Named entity recognition using an HMM-based chunk tagger,” in *proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 2002, pp. 473–480.
- [17] D. M. Bikel, S. Miller, R. Schwartz, and R. Weischedel, “Nymble: A High-performance Learning Name-finder,” in *Proceedings of the Fifth Conference on Applied Natural Language Processing*, Stroudsburg, PA, USA, 1997, pp. 194–201.
- [18] A. L. Berger, V. J. D. Pietra, and S. A. D. Pietra, “A Maximum Entropy Approach to Natural Language Processing,” *Comput. Linguist.*, vol. 22, no. 1, pp. 39–71, Mar. 1996.
- [19] M. Collins, “Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with Perceptron Algorithms,” in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, Stroudsburg, PA, USA, 2002, pp. 1–8.
- [20] J. Lafferty, A. McCallum, and F. Pereira, “Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data,” *Departmental Papers (CIS)*, Jun. 2001.
- [21] G. Bello-Orgaz, J. J. Jung, and D. Camacho, “Social big data: Recent achievements and new challenges,” *Information Fusion*, vol. 28, pp. 45–59, 2016.
- [22] A. Gandomi and M. Haider, “Beyond the hype: Big data concepts, methods, and analytics,” *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015.
- [23] M. Stonebraker, U. Çetintemel, and S. Zdonik, “The 8 Requirements of Real-time Stream Processing,” *SIGMOD Rec.*, vol. 34, no. 4, pp. 42–47, Dec. 2005.
- [24] A. Toshniwal *et al.*, “Storm@Twitter,” in *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, New York, NY, USA, 2014, pp. 147–156.
- [25] H. Gokavarapu, “Exploring Cassandra and HBase with BigTable Model,” *Indiana University Bloomington*.
- [26] Z. Hasani, M. Kon-Popovska, and G. Velinov, “Lambda architecture for real time big data analytic,” *ICT Innovations*, pp. 133–143, 2014.
- [27] J. M. Pierre, “Mining knowledge from text collections using automatically generated metadata,” in *International Conference on Practical Aspects of Knowledge Management*, 2002, pp. 537–548.
- [28] E. S. Ristad and P. N. Yianilos, “Learning string-edit distance,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 5, pp. 522–532, May 1998.

- [29] A. B. M. Moniruzzaman and S. A. Hossain, “NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison,” *arXiv:1307.0191 [cs]*, Jun. 2013.
- [30] S. Suhothayan, K. Gajasinghe, I. Loku Narangoda, S. Chaturanga, S. Perera, and V. Nanayakkara, “Siddhi: A Second Look at Complex Event Processing Architectures,” in *Proceedings of the 2011 ACM Workshop on Gateway Computing Environments*, New York, NY, USA, 2011, pp. 43–50.
- [31] J. Z. Pan, “Resource Description Framework,” in *Handbook on Ontologies*, Springer, Berlin, Heidelberg, 2009, pp. 71–90.
- [32] A. Jena, “Apache Jena Fuseki,” *The Apache Software Foundation*, 2014.
- [33] O. Hartig, C. Bizer, and J.-C. Freytag, “Executing SPARQL Queries over the Web of Linked Data,” in *The Semantic Web - ISWC 2009*, 2009, pp. 293–309.
- [34] R. He and J. McAuley, “Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering,” in *proceedings of the 25th international conference on world wide web*, 2016, pp. 507–517.
- [35] P. Petrovski, A. Primpeli, R. Meusel, and C. Bizer, “The WDC gold standards for product feature extraction and product matching,” in *International Conference on Electronic Commerce and Web Technologies*, 2016, pp. 73–86.
- [36] J. McAuley and A. Yang, “Addressing Complex and Subjective Product-Related Queries with Customer Reviews,” in *Proceedings of the 25th International Conference on World Wide Web*, Republic and Canton of Geneva, Switzerland, 2016, pp. 625–635.
- [37] P. Petrovski, A. Primpeli, R. Meusel, and C. Bizer, “The WDC Gold Standards for Product Feature Extraction and Product Matching,” in *E-Commerce and Web Technologies*, 2016, pp. 73–86.
- [38] L. Derczynski, A. Ritter, and S. Clark, *Twitter Part-of-Speech Tagging for All: Overcoming Sparse and Noisy Data*. .
- [39] G. A. Miller, “WordNet: a lexical database for English,” *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.
- [40] W. B. Cavnar and J. M. Trenkle, “Ngram-based text categorization,” in *In Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*, 1994, pp. 161–175.
- [41] B. Carpenter, “LingPipe for 99.99% recall of gene mentions,” in *Proceedings of the Second BioCreative Challenge Evaluation Workshop*, 2007, vol. 23, pp. 307–309.
- [42] B. Hollerit, M. Kröll, and M. Strohmaier, “Towards Linking Buyers and Sellers: Detecting Commercial Intent on Twitter,” in *Proceedings of the 22Nd International Conference on World Wide Web*, New York, NY, USA, 2013, pp. 629–632.
- [43] A. Jain, *Mastering Apache Storm: Real-time Big Data Streaming Using Kafka, Hbase and Redis*. Packt Publishing, 2017.

- [44] A. Chebotko, A. Kashlev, and S. Lu, “A Big Data Modeling Methodology for Apache Cassandra,” in *2015 IEEE International Congress on Big Data*, 2015, pp. 238–245.
- [45] N. Neeraj, *Mastering Apache Cassandra*. Packt Publishing Ltd, 2013.
- [46] W. Koehrsen, “Beyond Accuracy: Precision and Recall,” *Towards Data Science*, 03-Mar-2018. [Online]. Available: <https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c>. [Accessed: 10-Jun-2019].