

References

- [1] A. Das and B. Gambäck, “Code-mixing in social media text: the last language identification frontier?,” 2015.
- [2] P. McNamee, “Language Identification: A Solved Problem Suitable for Undergraduate Instruction,” *J Comput Sci Coll*, vol. 20, no. 3, pp. 94–101, Feb. 2005.
- [3] U. Barman, A. Das, J. Wagner, and J. Foster, “Code mixing: A challenge for language identification in the language of social media,” in *Proceedings of the first workshop on computational approaches to code switching*, 2014, pp. 13–23.
- [4] A. Das and B. Gambäck, “Identifying languages at the word level in code-mixed indian social media text,” 2014.
- [5] D. Crystal, *Language and the Internet*. Cambridge [u.a.: Cambridge Univ. Press, 2008.
- [6] T. Hidayat, *An analysis of code switching used by facebookers*. Sekolah Tinggi Keguruan dan Ilmu Pendidikan (STKIP) Siliwangi Bandung, 2008.
- [7] P. Muysken, “code-switching and grammatical theory,” p. 23.
- [8] P. Muysken, *Bilingual speech: a typology of code-mixing*, Nachdr. Cambridge: Cambridge Univ. Press, 2002.
- [9] “Sociolinguistics by Bernard Spolsky,” p. 5.
- [10] L. S. Kia, X. Cheng, T. K. Yee, and C. W. Ling, “Code-Mixing of English in the Entertainment News of Chinese Newspapers in Malaysia,” *Int. J. Engl. Linguist.*, vol. 1, no. 1, Mar. 2011.
- [11] S. Ajita and K. Amit, “POS Tagging of Hindi-English Code Mixed Text from Social Media.”
- [12] A. Ehsan and S. A. Aziz, “CODE-MIXING IN URDU NEWS OF A PRIVATE PAKISTANI CHANNEL: A CASE STUDY,” vol. 5, no. 1, p. 10, 2014.
- [13] T. Hidayat, *An analysis of code switching used by facebookers*. Sekolah Tinggi Keguruan dan Ilmu Pendidikan (STKIP) Siliwangi Bandung, 2008.
- [14] S. Ghosh, S. Ghosh, and D. Das, “Complexity Metric for Code-Mixed Social Media Text,” *ArXiv170701183 Cs*, Jul. 2017.
- [15] F. Fahmee and M. F. Yong, “Language Choice in Online Written Communication among Maldivian Professionals,” *3L Lang. Linguist. Lit.*, vol. 22, no. 2, 2016.
- [16] A. K. Joshi, “processing of sentences with intra-sentential code-switching,” in *Coling 1982: Proceedings of the Ninth International Conference on Computational Linguistics*, 1982.
- [17] E. M. GOLD, “Language Identification in the Limit,” p. 28.
- [18] P. V. Veena, M. Anand Kumar, and K. P. Soman, “Character Embedding for Language Identification in Hindi-English Code-mixed Social Media Text,” *Comput. Sist.*, vol. 22, no. 1, Mar. 2018.

- [19] D. Nguyen and A. S. Dođruöz, “Word Level Language Identification in Online Multilingual Communication,” in Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, Washington, USA, 2013, pp. 857–862.
- [20] C. Lignos and B. E. Al, Toward Web-scale Analysis of Codeswitching.
- [21] R. Řehůrek and M. Kolkus, “Language identification on the web: Extending the dictionary method,” in International Conference on Intelligent Text Processing and Computational Linguistics, 2009, pp. 357–368.
- [22] W. B. Cavnar and J. M. Trenkle, “N-gram-based text categorization,” *Ann Arbor Mi*, vol. 48113, no. 2, pp. 161–175, 1994.
- [23] T. Joachims, “Svmlight: Support vector machine,” SVM-Light Support Vector Mach. Httpsvmlight Joachims Org Univ. Dortmund., vol. 19, no. 4, 1999.
- [24] T. Baldwin and M. Lui, “Language identification: The long and the short of the matter,” in Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010, pp. 229–237.
- [25] B. King and S. Abney, “Labeling the Languages of Words in Mixed-Language Documents using Weakly Supervised Methods,” in Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Atlanta, Georgia, 2013, pp. 1110–1119.
- [26] Y. Vyas, S. Gella, J. Sharma, K. Bali, and M. Choudhury, “POS Tagging of English-Hindi Code-Mixed Social Media Content,” in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 2014, pp. 974–979.
- [27] M. Piergallini, R. Shirvani, G. S. Gautam, and M. Chouikha, “Word-Level Language Identification and Predicting Codeswitching Points in Swahili-English Language Data,” in Proceedings of the Second Workshop on Computational Approaches to Code Switching, Austin, Texas, 2016, pp. 21–29
- [28] S. Carter, W. Weerkamp, and M. Tsagkias, “Microblog language identification: overcoming the limitations of short, unedited and idiomatic text,” *Lang. Resour. Eval.*, vol. 47, no. 1, pp. 195–215, Mar. 2013.
- [29] T. Solorio et al., “Overview for the First Shared Task on Language Identification in Code-Switched Data,” in Proceedings of the First Workshop on Computational Approaches to Code Switching, Doha, Qatar, 2014, pp. 62–72.
- [30] G. Chittaranjan, Y. Vyas, K. Bali, and M. Choudhury, “Word-level Language Identification using CRF: Code-switching Shared Task Report of MSR India System,” in Proceedings of the First Workshop on Computational Approaches to Code Switching, Doha, Qatar, 2014, pp. 73–79.
- [31] R. Kumar, A. Kumar, K. P. Soman, and K. P. Soman, “AmritaCEN_NLP@ FIRE 2015 Language Identification for Indian Languages in Social Media Text,” 2015.
- [32] S. Deepu and R. Pethuru, “A Framework for Text Analytics using the Bag of Words (BoW) Model for Prediction,” *Int. J. Adv. Netw. Appl. IJANA*.
- [33] E. Charniak, Introduction to Artificial Intelligence. Pearson Education India, 1985

- [34] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: an introduction," *J. Am. Med. Inform. Assoc.*, vol. 18, no. 5, pp. 544–551, Sep. 2011.
- [35] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, Inc., 2009.
- [36] W.-L. Chao, "Machine Learning Tutorial," p. 56.
- [37] S. R. Kalmegh, "Comparative Analysis of WEKA Data Mining Algorithm RandomForest, RandomTree and LADTree for Classification of Indigenous News Data," vol. 5, no. 1, p. 11, 2015.
- [38] F. Pérez-Cruz, P. L. Alarcón-Diana, A. Navia-Vázquez, and A. Artés-Rodríguez, "Fast Training of Support Vector Classifiers," in *Advances in Neural Information Processing Systems 13*, T. K. Leen, T. G. Dietterich, and V. Tresp, Eds. MIT Press, 2001, pp. 734–740.
- [39] S. le Cessie and J. C. van Houwelingen, "Ridge Estimators in Logistic Regression," *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 41, no. 1, pp. 191–201, Mar. 1992.
- [40] George H. John, Pat Langley: Estimating Continuous Distributions in Bayesian Classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, 338-345, 1995.
- [41] I. H. Witten, E. Frank, L. Trigg, M. Hall, G. Holmes, and S. J. Cunningham, "Weka: Practical Machine Learning Tools and Techniques with Java Implementations," p. 5.
- [42] S. L. Salzberg, "C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993," *Mach. Learn.*, vol. 16, no. 3, pp. 235–240, Sep. 1994.
- [43] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
- [44] G. Nelson, "Guy Aston and Lou Burnard, *The BNC handbook: exploring the British National Corpus with SARA*. Edinburgh Textbooks in Empirical Linguistics. Edinburgh: Edinburgh University Press, 1998. Pp. 256. Hardback £43.50, ISBN 0 7486 1054 5; paperback £16.50, ISBN 0 7486 1055 3," *Engl. Lang. Linguist.*, vol. 6, no. 1, pp. 197–221, May 2002.
- [45] B. Han, P. Cook, and T. Baldwin, "Automatically constructing a normalisation dictionary for microblogs," in *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*, 2012, pp. 421–432.