

# References

- [1] G. I. Winata, A. F. Aji, Z. X. Yong, and T. Solorio, “The decades progress on code-switching research in NLP: A systematic survey on trends and challenges,” in *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, A. Rogers, J. L. Boyd-Graber, and N. Okazaki, Eds. Association for Computational Linguistics, 2023, pp. 2936–2978. [Online]. Available: <https://doi.org/10.18653/v1/2023.findings-acl.185>
- [2] D. M. Boyd and N. B. Ellison, “Social network sites: Definition, history, and scholarship,” *J. Comput. Mediat. Commun.*, 2007, vol. 13, no. 1, pp. 210–230. [Online]. Available: <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- [3] C. Myers-Scotton, “Common and Uncommon Ground: Social and Structural Factors in Codeswitching,” *Language in Society*, 1993, vol. 22, no. 4, pp. 475–503. [Online]. Available: <http://www.jstor.org/stable/4168471>
- [4] P. Rowlett, ““franglais”, concise encyclopedia of languages of the world,” pp. 425–426.
- [5] S. Rijhwani, R. Sequiera, M. Choudhury, K. Bali, and C. S. Maddila, “Estimating code-switching on Twitter with a novel generalized word-level language detection technique,” in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 1971–1982. [Online]. Available: <https://www.aclweb.org/anthology/P17-1180>
- [6] K. Bali, J. Sharma, M. Choudhury, and Y. Vyas, ““I am borrowing ya mixing ?” an analysis of English-Hindi code mixing in Facebook,” in *Proceedings of the First Workshop on Computational Approaches to Code Switching*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 116–126. [Online]. Available: <https://www.aclweb.org/anthology/W14-3914>
- [7] P. Agarwal, A. Sharma, J. Grover, M. Sikka, K. Rudra, and M. Choudhury, “I may talk in english but gaali toh hindi mein hi denge : A study of english-hindi code-switching

- and swearing pattern on social networks,” in *2017 9th International Conference on Communication Systems and Networks (COMSNETS)*, 2017, pp. 554–557.
- [8] S. Sitaram, K. R. Chandu, S. K. Rallabandi, and A. W. Black, “A survey of code-switched speech and language processing,” *CoRR*, 2019, vol. abs/1904.00784. [Online]. Available: <http://arxiv.org/abs/1904.00784>
- [9] A. I. Ansaldo, K. Marcotte, L. Scherer, and G. Raboyeau, “Language therapy and bilingual aphasia: Clinical implications of psycholinguistic and neuroimaging research,” *Journal of Neurolinguistics*, 2008, vol. 21, no. 6, pp. 539–557, acquisition, Processing and Loss of L2: Functional, cognitive and neural perspectives. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0911604408000146>
- [10] K. Chakma and A. Das, “CMIR: A corpus for evaluation of code mixed information retrieval of hindi-english tweets,” *Computación y Sistemas*, 2016, vol. 20, no. 3, pp. 425–434. [Online]. Available: <https://doi.org/10.13053/cys-20-3-2459>
- [11] B. Gambäck, “On measuring the complexity of code-mixing,” vol. 2014, 2014, pp. 1–7.
- [12] R. Barnett, E. Codó, E. Eppler, M. Forcadell, P. Gardner-Chloros, R. V. Hout, M. Moyer, M. C. Torras, M. T. Turell, M. Sebba, M. Starren, and S. Wensing, “The lides coding manual: A document for preparing and analyzing language interaction data version 1.1—july, 1999,” *International Journal of Bilingualism*, 2000, vol. 4, no. 2, pp. 131–132. [Online]. Available: <https://doi.org/10.1177/13670069000040020101>
- [13] G. A. Guzmán, J. Serigos, B. E. Bullock, and A. J. Toribio, “Simple tools for exploring variation in code-switching for linguists,” in *Proceedings of the Second Workshop on Computational Approaches to Code Switching@EMNLP 2016, Austin, Texas, USA, November 1, 2016*, M. T. Diab, P. Fung, M. Ghoneim, J. Hirschberg, and T. Solorio, Eds. Association for Computational Linguistics, 2016, pp. 12–20. [Online]. Available: <https://doi.org/10.18653/v1/W16-5802>
- [14] J. Zobel and P. Dart, “Phonetic string matching: Lessons from information retrieval,” in *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR ’96. New York, NY, USA: Association for Computing Machinery, 1996, p. 166–172. [Online]. Available: <https://doi.org/10.1145/243199.243258>
- [15] R. C. Russell, “The soundex coding system,” U.S. Patent 1 261 167, April. 1918.
- [16] T. Gadd, “‘fishing fore werds’: phonetic retrieval of written text in information systems,” *Program: electronic library and information systems*, 1988, vol. 22, no. 3, pp. 222–237. [Online]. Available: <https://doi.org/10.1108/eb046999>

- [17] T. N. Gadd, “Phonix: The algorithm,” *Program: electronic library and information systems*, 1990, vol. 24, no. 4, pp. 363–366. [Online]. Available: <https://doi.org/10.1108/eb047069>
- [18] D. K. Prabhakar, S. Pal, and C. Kumar, “Query expansion for transliterated text retrieval,” *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, jul 2021, vol. 20, no. 4. [Online]. Available: <https://doi.org/10.1145/3447649>
- [19] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval*. Cambridge University Press, 2008. [Online]. Available: <https://nlp.stanford.edu/IR-book/pdf/irbookprint.pdf>
- [20] G. Amati and C. J. Van Rijsbergen, “Probabilistic models of information retrieval based on measuring the divergence from randomness,” *ACM Trans. Inf. Syst.*, oct 2002, vol. 20, no. 4, p. 357–389. [Online]. Available: <https://doi.org/10.1145/582415.582416>
- [21] G. Amati and C. J. V. Rijsbergen, “Term frequency normalization via pareto distributions,” in *Advances in Information Retrieval, 24th BCS-IRSG European Colloquium on IR Research Glasgow, UK, March 25-27, 2002 Proceedings*, ser. Lecture Notes in Computer Science, F. Crestani, M. A. Girolami, and C. J. van Rijsbergen, Eds., vol. 2291. Springer, 2002, pp. 183–192. [Online]. Available: [https://doi.org/10.1007/3-540-45886-7\\_13](https://doi.org/10.1007/3-540-45886-7_13)
- [22] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” in *Proceedings of the Eighteenth International Conference on Machine Learning*, ser. ICML ’01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, p. 282–289.
- [23] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” in *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2013. [Online]. Available: <http://arxiv.org/abs/1301.3781>
- [24] T. Mikolov, E. Grave, P. Bojanowski, C. Puhersch, and A. Joulin, “Advances in pre-training distributed word representations,” in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA), 2018. [Online]. Available: <https://aclanthology.org/L18-1008>
- [25] J. Pennington, R. Socher, and C. Manning, “GloVe: Global vectors for word representation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational

- Linguistics, 2014, pp. 1532–1543. [Online]. Available: <https://aclanthology.org/D14-1162>
- [26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, Eds., 2017, pp. 5998–6008. [Online]. Available: <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>
- [27] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, 2019, pp. 4171–4186. [Online]. Available: <https://aclanthology.org/N19-1423>
- [28] S. Mustonen, “Multiple discriminant analysis in linguistic problems,” *Statistical Methods in Linguistics*, 1965, vol. 4, pp. 37–44.
- [29] M. D. Rau, “Language identification by statistical analysis,” Ph.D. dissertation, Naval Postgraduate School, 1974. [Online]. Available: <https://core.ac.uk/download/pdf/36710344.pdf>
- [30] P. Henrich, “Language identification for the automatic grapheme-to-phoneme conversion of foreign words in a german text-to-speech system,” in *First European Conference on Speech Communication and Technology, EUROSPEECH 1989, Paris, France, September 27-29, 1989*. ISCA, 1989, pp. 2220–2223. [Online]. Available: <https://doi.org/10.21437/Eurospeech.1989-197>
- [31] T. Vitale, “An algorithm for high accuracy name pronunciation by parametric speech synthesizer,” *Comput. Linguistics*, 1991, vol. 17, no. 3, pp. 257–276.
- [32] C. Souter, G. Churcher, J. Hayes, J. Hughes, and S. Johnson, “Natural language identification using corpus-based models,” *HERMES-Journal of Language and Communication in Business*, 1994, no. 13, pp. 183–203.
- [33] E. Giguët, “Categorization according to language: A step toward combining linguistic knowledge and statistic learning,” in *International Workshop of Parsing Technologies (IWPT’95)*, 1995.

- [34] G. Kikui, “Identifying the coding system and language of on-line documents on the internet,” in *16th International Conference on Computational Linguistics, Proceedings of the Conference, COLING 1996, Center for Sprogteknologi, Copenhagen, Denmark, August 5-9, 1996*, 1996, pp. 652–657. [Online]. Available: <https://aclanthology.org/C96-2110/>
- [35] L. A. Mather, “A linear algebra approach to language identification,” in *International Workshop on Principles of Digital Document Processing*. Springer, 1998, pp. 92–103.
- [36] D. Elworthy, “Language identification with confidence limits,” *CoRR*, 1999, vol. cs.CL/9907010. [Online]. Available: <https://arxiv.org/abs/cs/9907010>
- [37] W. B. Cavnar and J. M. Trenkle, “N-gram-based text categorization,” in *In Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*, 1994, pp. 161–175.
- [38] D. Nguyen and A. S. Doğruöz, “Word level language identification in online multilingual communication,” in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*. ACL, 2013, pp. 857–862. [Online]. Available: <https://aclanthology.org/D13-1084/>
- [39] 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@EMNLP 2014, Doha, Qatar, October 25, 2014*, M. T. Diab, J. Hirschberg, P. Fung, and T. Solorio, Eds. Association for Computational Linguistics, 2014, pp. 13–23. [Online]. Available: <https://doi.org/10.3115/v1/W14-3902>
- [40] A. Das and B. Gambäck, “Identifying languages at the word level in code-mixed indian social media text,” in *Proceedings of the 11th International Conference on Natural Language Processing, ICON 2014, Goa, India, December 18-21, 2014*, D. M. Sharma, R. Sangal, and J. D. Pawar, Eds. NLP Association of India, 2014, pp. 378–387. [Online]. Available: <https://aclanthology.org/W14-5152/>
- [41] 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: Association for Computational Linguistics, 2014, pp. 974–979. [Online]. Available: <https://aclanthology.org/D14-1105>
- [42] M. X. Xia, “Codeswitching language identification using subword information enriched word vectors,” in *Proceedings of the Second Workshop on Computational Approaches to*

- Code Switching*. Austin, Texas: Association for Computational Linguistics, 2016, pp. 132–136. [Online]. Available: <https://aclanthology.org/W16-5818>
- [43] 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@EMNLP 2014, Doha, Qatar, October 25, 2014*, M. T. Diab, J. Hirschberg, P. Fung, and T. Solorio, Eds. Association for Computational Linguistics, 2014, pp. 73–79. [Online]. Available: <https://doi.org/10.3115/v1/W14-3908>
- [44] P. V. Veena, M. A. Kumar, and K. P. Soman, “Character embedding for language identification in hindi-english code-mixed social media text,” *Computación y Sistemas*, 2018, vol. 22, no. 1. [Online]. Available: <https://doi.org/10.13053/cys-22-1-2775>
- [45] H. Jhamtani, S. K. Bhogi, and V. Raychoudhury, “Word-level language identification in bi-lingual code-switched texts,” in *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing*. Phuket, Thailand: Department of Linguistics, Chulalongkorn University, 2014, pp. 348–357. [Online]. Available: <https://aclanthology.org/Y14-1041>
- [46] M. Z. Ansari, S. Khan, T. Amani, A. Hamid, and S. R. A. Rizvi, “Analysis of part of speech tags in language identification of code-mixed text,” in *Proceedings of International Conference on Advancements in Computing and Management (ICACM)*, 2019. [Online]. Available: <https://ssrn.com/abstract=3462976>
- [47] S. Thara and P. Poornachandran, “Code-mixing: A brief survey,” in *2018 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2018, Bangalore, India, September 19-22, 2018*. IEEE, 2018, pp. 2382–2388. [Online]. Available: <https://doi.org/10.1109/ICACCI.2018.8554413>
- [48] T. Jauhiainen, M. Lui, M. Zampieri, T. Baldwin, and K. Lindén, “Automatic language identification in texts: A survey,” *J. Artif. Int. Res.*, 2019, vol. 65, no. 1, p. 675–682. [Online]. Available: <https://doi.org/10.1613/jair.1.11675>
- [49] N. Sarma, S. R. Singh, and D. Goswami, “Switchnet: Learning to switch for word-level language identification in code-mixed social media text,” *Natural Language Engineering*, 2022, vol. 28, no. 3, pp. 337–359. [Online]. Available: <https://doi.org/10.1017/S1351324921000115>
- [50] G. Aguilar, S. Kar, and T. Solorio, “Lince: A centralized benchmark for linguistic code-switching evaluation,” in *Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020*, N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara,

- B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, Eds. European Language Resources Association, 2020, pp. 1803–1813. [Online]. Available: <https://aclanthology.org/2020.lrec-1.223/>
- [51] G. Molina, F. AlGhamdi, M. Ghoneim, A. Hawwari, N. Rey-Villamizar, M. Diab, and T. Solorio, “Overview for the second shared task on language identification in code-switched data,” in *Proceedings of the Second Workshop on Computational Approaches to Code Switching*. Austin, Texas: Association for Computational Linguistics, 2016, pp. 40–49. [Online]. Available: <https://aclanthology.org/W16-5805>
- [52] T. Solorio, E. Blair, S. Maharjan, S. Bethard, M. Diab, M. Ghoneim, A. Hawwari, F. AlGhamdi, J. Hirschberg, A. Chang, and P. Fung, “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: Association for Computational Linguistics, 2014, pp. 62–72. [Online]. Available: <https://aclanthology.org/W14-3907>
- [53] D. Mave, S. Maharjan, and T. Solorio, “Language identification and analysis of code-switched social media text,” in *Proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching*. Melbourne, Australia: Association for Computational Linguistics, 2018, pp. 51–61. [Online]. Available: <https://aclanthology.org/W18-3206>
- [54] S. Khanuja, S. Dandapat, A. Srinivasan, S. Sitaram, and M. Choudhury, “Gluecos: An evaluation benchmark for code-switched NLP,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, D. Jurafsky, J. Chai, N. Schluter, and J. R. Tetreault, Eds. Association for Computational Linguistics, 2020, pp. 3575–3585. [Online]. Available: <https://doi.org/10.18653/v1/2020.acl-main.329>
- [55] F. Balouchzahi, S. Butt, A. Hegde, N. Ashraf, H. Shashirekha, G. Sidorov, and A. Gelbukh, “Overview of CoLI-kanglish: Word level language identification in code-mixed Kannada-English texts at ICON 2022,” in *Proceedings of the 19th International Conference on Natural Language Processing (ICON): Shared Task on Word Level Language Identification in Code-mixed Kannada-English Texts*, B. R. Chakravarthi, A. Murugappan, D. Chinnappa, A. Hane, P. K. Kumeresan, and R. Ponnusamy, Eds. IIT Delhi, New Delhi, India: Association for Computational Linguistics, Dec. 2022, pp. 38–45. [Online]. Available: <https://aclanthology.org/2022.icon-wlli.8>

- [56] S. H. Lakshmaiah, F. Balouchzahi, M. D. Anusha, and G. Sidorov, “Coli-machine learning approaches for code-mixed language identification at the word level in kannada-english texts,” *Acta Polytechnica Hungarica*, 2022, vol. 19, no. 10, pp. 123–141.
- [57] A. Hegde, F. Balouchzahi, S. Coelho, S. H. L, and H. A. Nayel, “Coli@fire2023: Findings of word-level language identification in code-mixed tulu text,” in *Proceedings of the 15th Annual Meeting of the Forum for Information Retrieval Evaluation*, ser. FIRE ’23. New York, NY, USA: Association for Computing Machinery, 2024, p. 25–26. [Online]. Available: <https://doi.org/10.1145/3632754.3633075>
- [58] R. Nayak and R. Joshi, “L3Cube-HingCorpus and HingBERT: A code mixed Hindi-English dataset and BERT language models,” in *Proceedings of the WILDRE-6 Workshop within the 13th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, Jun. 2022, pp. 7–12. [Online]. Available: <https://aclanthology.org/2022.wildre-1.2>
- [59] N. J. Kalita, A. G. Agarwala, and J. Das, “Word level language identification on code-mixed english-bodo text,” *IOP Conference Series: Materials Science and Engineering*, jan 2021, vol. 1020, no. 1, p. 012027. [Online]. Available: <https://dx.doi.org/10.1088/1757-899X/1020/1/012027>
- [60] M. Yasir, L. Chen, A. Khatoun, M. A. Malik, F. Abid, and A. M. Khalil, “Mixed script identification using automated dnn hyperparameter optimization,” *Intell. Neuroscience*, jan 2021, vol. 2021. [Online]. Available: <https://doi.org/10.1155/2021/8415333>
- [61] A. Dutta, “Word-level language identification using subword embeddings for code-mixed Bangla-English social media data,” in *Proceedings of the Workshop on Dataset Creation for Lower-Resourced Languages within the 13th Language Resources and Evaluation Conference*, J. Sälevä and C. Lignos, Eds. Marseille, France: European Language Resources Association, Jun. 2022, pp. 76–82. [Online]. Available: <https://aclanthology.org/2022.dclrl-1.10>
- [62] P. Rani, J. P. McCrae, and T. Franssen, “MHE: Code-mixed corpora for similar language identification,” in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, and S. Piperidis, Eds. Marseille, France: European Language Resources Association, Jun. 2022, pp. 3425–3433. [Online]. Available: <https://aclanthology.org/2022.lrec-1.366>
- [63] R. Joshi and R. Joshi, “Evaluating input representation for language identification in hindi-english code mixed text,” *CoRR*, 2020, vol. abs/2011.11263. [Online]. Available: <https://arxiv.org/abs/2011.11263>

- [64] A. Jamatia, A. Das, and B. Gambäck, “Deep learning-based language identification in english-hindi-bengali code-mixed social media corpora,” *J. Intell. Syst.*, 2019, vol. 28, no. 3, pp. 399–408. [Online]. Available: <https://doi.org/10.1515/jisys-2017-0440>
- [65] S. Sharma, S. PYKL, and R. C. Balabantaray, “Text normalization of code mix and sentiment analysis,” in *2015 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2015, Kochi, India, August 10-13, 2015*. IEEE, 2015, pp. 1468–1473. [Online]. Available: <https://doi.org/10.1109/ICACCI.2015.7275819>
- [66] A. Joshi, A. Prabhu, M. Shrivastava, and V. Varma, “Towards sub-word level compositions for sentiment analysis of Hindi-English code mixed text,” in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, Y. Matsumoto and R. Prasad, Eds. Osaka, Japan: The COLING 2016 Organizing Committee, Dec. 2016, pp. 2482–2491. [Online]. Available: <https://aclanthology.org/C16-1234>
- [67] S. Mandal and D. Das, “Analyzing roles of classifiers and code-mixed factors for sentiment identification,” *CoRR*, 2018, vol. abs/1801.02581. [Online]. Available: <http://arxiv.org/abs/1801.02581>
- [68] P. Patwa, G. Aguilar, S. Kar, S. Pandey, S. PYKL, B. Gambäck, T. Chakraborty, T. Solorio, and A. Das, “SemEval-2020 task 9: Overview of sentiment analysis of code-mixed tweets,” in *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, A. Herbelot, X. Zhu, A. Palmer, N. Schneider, J. May, and E. Shutova, Eds. Barcelona (online): International Committee for Computational Linguistics, Dec. 2020, pp. 774–790. [Online]. Available: <https://aclanthology.org/2020.semeval-1.100>
- [69] D. Vilares, M. A. Alonso, and C. Gómez-Rodríguez, “Sentiment analysis on monolingual, multilingual and code-switching Twitter corpora,” in *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, A. Balahur, E. Van der Goot, P. Vossen, and A. Montoyo, Eds. Lisboa, Portugal: Association for Computational Linguistics, Sep. 2015, pp. 2–8. [Online]. Available: <https://aclanthology.org/W15-2902>
- [70] D. Vilares, M. A. Alonso, and C. Gómez-Rodríguez, “EN-ES-CS: an english-spanish code-switching twitter corpus for multilingual sentiment analysis,” in *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016*, N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odiijk, and S. Piperidis, Eds. European Language Resources Association (ELRA), 2016. [Online]. Available: <http://www.lrec-conf.org/proceedings/lrec2016/summaries/43.html>

- [71] D. Vilares, M. A. Alonso, and C. Gómez-Rodríguez, “Supervised sentiment analysis in multilingual environments,” *Information Processing & Management*, 2017, vol. 53, no. 3, pp. 595–607. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306457316302540>
- [72] S. Lee and Z. Wang, “Emotion in code-switching texts: Corpus construction and analysis,” in *Proceedings of the Eighth SIGHAN Workshop on Chinese Language Processing*, L.-C. Yu, Z. Sui, Y. Zhang, and V. Ng, Eds. Beijing, China: Association for Computational Linguistics, Jul. 2015, pp. 91–99. [Online]. Available: <https://aclanthology.org/W15-3116>
- [73] Z. Wang, S. Lee, S. Li, and G. Zhou, “Emotion detection in code-switching texts via bilingual and sentimental information,” in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, C. Zong and M. Strube, Eds. Beijing, China: Association for Computational Linguistics, Jul. 2015, pp. 763–768. [Online]. Available: <https://aclanthology.org/P15-2125>
- [74] G. Aguilar, S. Kar, and T. Solorio, “LinCE: A centralized benchmark for linguistic code-switching evaluation,” in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, Eds. Marseille, France: European Language Resources Association, May 2020, pp. 1803–1813. [Online]. Available: <https://aclanthology.org/2020.lrec-1.223>
- [75] A. R. Appidi, V. K. Srirangam, D. Suhas, and M. Shrivastava, “Creation of corpus and analysis in code-mixed Kannada-English Twitter data for emotion prediction,” in *Proceedings of the 28th International Conference on Computational Linguistics*, D. Scott, N. Bel, and C. Zong, Eds. Barcelona, Spain (Online): International Committee on Computational Linguistics, Dec. 2020, pp. 6703–6709. [Online]. Available: <https://aclanthology.org/2020.coling-main.587>
- [76] S. Padmaja, S. Fatima, S. Bandu, M. Nikitha, and K. Prathyusha, “Sentiment extraction from bilingual code mixed social media text,” in *Data Engineering and Communication Technology*, K. S. Raju, R. Senkerik, S. P. Lanka, and V. Rajagopal, Eds. Singapore: Springer Nature Singapore, 2020, pp. 707–714.
- [77] S. S. V. Kusampudi, P. Sathineni, and R. Mamidi, “Sentiment analysis in code-mixed Telugu-English text with unsupervised data normalization,” in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, R. Mitkov and G. Angelova, Eds. Held Online: INCOMA Ltd., Sep. 2021, pp. 753–760. [Online]. Available: <https://aclanthology.org/2021.ranlp-1.86>

- [78] S. Chanda and S. Pal, “Irlab@iitbhu@dravidian-codemix-fire2020: Sentiment analysis for dravidian languages in code-mixed text,” in *Fire*, 2020. [Online]. Available: <https://api.semanticscholar.org/CorpusID:232314461>
- [79] S. Chanda, R. Singh, and S. Pal, “Is meta embedding better than pre-trained word embedding to perform sentiment analysis for dravidian languages in code-mixed text?” in *Fire*, 2021. [Online]. Available: <https://api.semanticscholar.org/CorpusID:251125487>
- [80] C. Nobata, J. Tetreault, A. Thomas, Y. Mehdad, and Y. Chang, “Abusive language detection in online user content,” in *Proceedings of the 25th International Conference on World Wide Web*, ser. WWW '16. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2016, p. 145–153. [Online]. Available: <https://doi.org/10.1145/2872427.2883062>
- [81] H. Mubarak, K. Darwish, and W. Magdy, “Abusive language detection on Arabic social media,” in *Proceedings of the First Workshop on Abusive Language Online*. Vancouver, BC, Canada: Association for Computational Linguistics, Aug. 2017, pp. 52–56. [Online]. Available: <https://www.aclweb.org/anthology/W17-3008>
- [82] R. Kumar, A. K. Ojha, S. Malmasi, and M. Zampieri, “Benchmarking aggression identification in social media,” in *Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018)*. Santa Fe, New Mexico, USA: Association for Computational Linguistics, Aug. 2018, pp. 1–11. [Online]. Available: <https://www.aclweb.org/anthology/W18-4401>
- [83] J.-M. Xu, K.-S. Jun, X. Zhu, and A. Bellmore, “Learning from bullying traces in social media,” in *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, ser. NAACL HLT '12. USA: Association for Computational Linguistics, 2012, p. 656–666.
- [84] M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, “Improving cyberbullying detection with user context,” in *Advances in Information Retrieval*, P. Serdyukov, P. Braslavski, S. O. Kuznetsov, J. Kamps, S. Rüger, E. Agichtein, I. Segalovich, and E. Yilmaz, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 693–696.
- [85] N. Djuric, J. Zhou, R. Morris, M. Grbovic, V. Radosavljevic, and N. Bhamidipati, “Hate speech detection with comment embeddings,” in *Proceedings of the 24th International Conference on World Wide Web*, ser. WWW '15 Companion. New York, NY, USA: Association for Computing Machinery, 2015, p. 29–30. [Online]. Available: <https://doi.org/10.1145/2740908.2742760>

- [86] T. Davidson, D. Warmley, M. W. Macy, and I. Weber, “Automated hate speech detection and the problem of offensive language,” *CoRR*, 2017, vol. abs/1703.04009. [Online]. Available: <http://arxiv.org/abs/1703.04009>
- [87] I. Kwok and Y. Wang, “Locate the hate: Detecting tweets against blacks,” in *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, ser. AAAI’13. AAAI Press, 2013, p. 1621–1622.
- [88] S. Malmasi and M. Zampieri, “Detecting hate speech in social media,” *CoRR*, 2017, vol. abs/1712.06427. [Online]. Available: <http://arxiv.org/abs/1712.06427>
- [89] T. Mandl, S. Modha, P. Majumder, D. Patel, M. Dave, C. Mandlia, and A. Patel, “Overview of the hasoc track at fire 2019: Hate speech and offensive content identification in indo-european languages,” in *Proceedings of the 11th Annual Meeting of the Forum for Information Retrieval Evaluation*, ser. FIRE ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 14–17. [Online]. Available: <https://doi.org/10.1145/3368567.3368584>
- [90] B. Ross, M. Rist, G. Carbonell, B. Cabrera, N. Kurowsky, and M. Wojatzki, “Measuring the reliability of hate speech annotations: The case of the european refugee crisis,” *CoRR*, 2017, vol. abs/1701.08118. [Online]. Available: <http://arxiv.org/abs/1701.08118>
- [91] H. Mubarak, A. Rashed, K. Darwish, Y. Samih, and A. Abdelali, “Arabic offensive language on twitter: Analysis and experiments,” *arXiv preprint arXiv:2004.02192*, 2020.
- [92] Z. Pitenis, M. Zampieri, and T. Ranasinghe, “Offensive Language Identification in Greek,” in *Proceedings of the 12th Language Resources and Evaluation Conference*. ELRA, 2020.
- [93] D. Fišer, T. Erjavec, and N. Ljubešić, “Legal framework, dataset and annotation schema for socially unacceptable online discourse practices in Slovene,” in *Proceedings of the First Workshop on Abusive Language Online*. Vancouver, BC, Canada: Association for Computational Linguistics, Aug. 2017, pp. 46–51. [Online]. Available: <https://www.aclweb.org/anthology/W17-3007>
- [94] H.-P. Su, Z.-J. Huang, H.-T. Chang, and C.-J. Lin, “Rephrasing profanity in Chinese text,” in *Proceedings of the First Workshop on Abusive Language Online*. Vancouver, BC, Canada: Association for Computational Linguistics, Aug. 2017, pp. 18–24. [Online]. Available: <https://www.aclweb.org/anthology/W17-3003>
- [95] A. Saroj, S. Chanda, and S. Pal, “IRlab@IITV at SemEval-2020 task 12: Multilingual offensive language identification in social media using SVM,” in *Proceedings of the Fourteenth Workshop on Semantic Evaluation*. Barcelona (online): International

- Committee for Computational Linguistics, Dec. 2020, pp. 2012–2016. [Online]. Available: <https://aclanthology.org/2020.semeval-1.265>
- [96] M. Wiegand, M. Siegel, and J. Ruppenhofer, “Overview of the germeval 2018 shared task on the identification of offensive language,” ser. Proceedings of GermEval 2018, 14th Conference on Natural Language Processing (KONVENS 2018), Vienna, Austria – September 21, 2018. Vienna, Austria: Austrian Academy of Sciences, 2018, pp. 1 – 10. [Online]. Available: <http://nbn-resolving.de/urn:nbn:de:bsz:mh39-84935>
- [97] B. R. Chakravarthi, R. Priyadharshini, N. Jose, A. Kumar M, T. Mandl, P. K. Kumaresan, R. Ponnusamy, H. R L, J. P. McCrae, and E. Sherly, “Findings of the shared task on offensive language identification in Tamil, Malayalam, and Kannada,” in *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Kyiv: Association for Computational Linguistics, Apr. 2021, pp. 133–145. [Online]. Available: <https://aclanthology.org/2021.dravidianlangtech-1.17>
- [98] S. Sai and Y. Sharma, “Towards offensive language identification for Dravidian languages,” in *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Kyiv: Association for Computational Linguistics, Apr. 2021, pp. 18–27. [Online]. Available: <https://aclanthology.org/2021.dravidianlangtech-1.3>
- [99] T. Ranasinghe, S. Gupte, M. Zampieri, and I. Nwogu, “WLV-RIT at hasoc-dravidian-codemix-fire2020: Offensive language identification in code-switched youtube comments,” *CoRR*, 2020, vol. abs/2011.00559. [Online]. Available: <https://arxiv.org/abs/2011.00559>
- [100] M. Zampieri, S. Malmasi, P. Nakov, S. Rosenthal, N. Farra, and R. Kumar, “Predicting the type and target of offensive posts in social media,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 1415–1420. [Online]. Available: <https://aclanthology.org/N19-1144>
- [101] A. Bagora, K. Shrestha, K. Maurya, and M. S. Desarkar, “Hostility detection in online hindi-english code-mixed conversations,” in *Proceedings of the 14th ACM Web Science Conference 2022*, ser. WebSci ’22. New York, NY, USA: Association for Computing Machinery, 2022, p. 390–400. [Online]. Available: <https://doi.org/10.1145/3501247.3531579>
- [102] H. Madhu, S. Satapara, S. Modha, T. Mandl, and P. Majumder, “Detecting offensive speech in conversational code-mixed dialogue on social media: A contextual dataset and benchmark experiments,” *Expert Systems with Applications*, 2023, vol.

- 215, p. 119342. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417422023600>
- [103] D. Ganguly, A. Bandyopadhyay, M. Mitra, and G. J. Jones, “Retrievability of code mixed microblogs,” in *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 973–976. [Online]. Available: <https://doi.org/10.1145/2911451.2914727>
- [104] P. Gupta, P. Rosso, and R. E. Banchs, “Encoding transliteration variation through dimensionality reduction: Fire shared task on transliterated search,” 2013. [Online]. Available: [https://www.isical.ac.in/~fire/wn/STTS/2013-translit\\_search-gupta-tu\\_valencia.pdf](https://www.isical.ac.in/~fire/wn/STTS/2013-translit_search-gupta-tu_valencia.pdf)
- [105] H. Joshi, A. Bhatt, and H. Patel, “Transliterated search using syllabification approach,” 2013. [Online]. Available: [https://www.isical.ac.in/~fire/wn/STTS/2013-translit\\_search-joshi-guj\\_univ.pdf](https://www.isical.ac.in/~fire/wn/STTS/2013-translit_search-joshi-guj_univ.pdf)
- [106] P. Pakray and P. Bhaskar, “Transliterated search system for indian languages,” 2013. [Online]. Available: [https://www.isical.ac.in/~fire/wn/STTS/2013-translit\\_search-pakray-ntnu-norway.pdf](https://www.isical.ac.in/~fire/wn/STTS/2013-translit_search-pakray-ntnu-norway.pdf)
- [107] A. Prakash and S. K. Saha, “A relevance feedback based approach for mixed script transliterated text search : Shared task report by bit mesra , india,” 2014. [Online]. Available: [https://www.isical.ac.in/~fire/working-notes/2014/MSR/FIRE2014\\_BIT.pdf](https://www.isical.ac.in/~fire/working-notes/2014/MSR/FIRE2014_BIT.pdf)
- [108] A. Mukherjee, A. Ravi, and K. Datta, “Mixed-script query labelling using supervised learning and ad hoc retrieval using sub word indexing,” in *Proceedings of the Forum for Information Retrieval Evaluation*, ser. FIRE '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 86–90. [Online]. Available: <https://doi.org/10.1145/2824864.2824873>
- [109] D. Ganguly, S. Pal, and G. J. F. Jones, “Dcu@fire-2014: Fuzzy queries with rule-based normalization for mixed script information retrieval,” in *Proceedings of the Forum for Information Retrieval Evaluation*, ser. FIRE '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 80–85. [Online]. Available: <https://doi.org/10.1145/2824864.2824888>
- [110] I. A. Bhat, V. Mujadia, A. Tammewar, R. A. Bhat, and M. Shrivastava, “Iiit-h system submission for fire2014 shared task on transliterated search,” in *Proceedings of the Forum for Information Retrieval Evaluation*, ser. FIRE '14. New York, NY,

- USA: Association for Computing Machinery, 2014, p. 48–53. [Online]. Available: <https://doi.org/10.1145/2824864.2824872>
- [111] N. Ghelani, S. K. Saha, and A. Prakash, “Mixed script ad hoc retrieval using back transliteration and phrase matching through bigram indexing: Shared task report by bit, mesra,” in *Post Proceedings of the Workshops at the 7th Forum for Information Retrieval Evaluation, Gandhinagar, India, December 4-6, 2015*, ser. CEUR Workshop Proceedings, P. Majumder, M. Mitra, M. Agrawal, and P. Mehta, Eds., vol. 1587. CEUR-WS.org, 2015, pp. 59–62. [Online]. Available: <https://ceur-ws.org/Vol-1587/T2-11.pdf>
- [112] D. Jain, “DA-IICT in FIRE 2015 shared task on mixed script information retrieval,” in *Post Proceedings of the Workshops at the 7th Forum for Information Retrieval Evaluation, Gandhinagar, India, December 4-6, 2015*, ser. CEUR Workshop Proceedings, P. Majumder, M. Mitra, M. Agrawal, and P. Mehta, Eds., vol. 1587. CEUR-WS.org, 2015, pp. 51–54. [Online]. Available: <http://ceur-ws.org/Vol-1587/T2-9.pdf>
- [113] D. K. Prabhakar and S. Pal, “Ism@fire-2015: Mixed script information retrieval,” in *Post Proceedings of the Workshops at the 7th Forum for Information Retrieval Evaluation, Gandhinagar, India, December 4-6, 2015*, ser. CEUR Workshop Proceedings, P. Majumder, M. Mitra, M. Agrawal, and P. Mehta, Eds., vol. 1587. CEUR-WS.org, 2015, pp. 55–58. [Online]. Available: <https://ceur-ws.org/Vol-1587/T2-10.pdf>
- [114] N. Londhe and R. K. Srihari, “Exploiting named entity mentions towards code mixed IR : Working notes for the UB system submission for msir@fire-2016,” in *Working notes of FIRE 2016 - Forum for Information Retrieval Evaluation, Kolkata, India, December 7-10, 2016*, ser. CEUR Workshop Proceedings, P. Majumder, M. Mitra, P. Mehta, J. Sankhavara, and K. Ghosh, Eds., vol. 1737. CEUR-WS.org, 2016, pp. 105–108. [Online]. Available: <https://ceur-ws.org/Vol-1737/T3-3.pdf>
- [115] H. B. B. Ganesh, M. A. Kumar, and K. P. Soman, “Distributional semantic representation for text classification and information retrieval,” in *Working notes of FIRE 2016 - Forum for Information Retrieval Evaluation, Kolkata, India, December 7-10, 2016*, ser. CEUR Workshop Proceedings, P. Majumder, M. Mitra, P. Mehta, J. Sankhavara, and K. Ghosh, Eds., vol. 1737. CEUR-WS.org, 2016, pp. 126–130. [Online]. Available: <https://ceur-ws.org/Vol-1737/T3-8.pdf>
- [116] S. Singh, M. A. Kumar, and K. P. Soman, “Cen@amrita: Information retrieval on codemixed hindi-english tweets using vector space models,” in *Working notes of FIRE 2016 - Forum for Information Retrieval Evaluation, Kolkata, India, December*

- 7-10, 2016, ser. CEUR Workshop Proceedings, P. Majumder, M. Mitra, P. Mehta, J. Sankhavara, and K. Ghosh, Eds., vol. 1737. CEUR-WS.org, 2016, pp. 131–134. [Online]. Available: <https://ceur-ws.org/Vol-1737/T3-9.pdf>
- [117] S. Banerjee, K. Chakma, S. K. Naskar, A. Das, P. Rosso, S. Bandyopadhyay, and M. Choudhury, “Overview of the mixed script information retrieval (msir),” in *Proceedings of FIRE 2016*. FIRE, December 2016. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/overview-mixed-script-information-retrieval-msir/>
- [118] N. Jose, B. R. Chakravarthi, S. Suryawanshi, E. Sherly, and J. P. McCrae, “A survey of current datasets for code-switching research,” in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2020, pp. 136–141.
- [119] B. G. Patra, D. Das, and A. Das, “Sentiment analysis of code-mixed indian languages: An overview of sail\_code-mixed shared task @icon-2017,” *CoRR*, 2018, vol. abs/1803.06745. [Online]. Available: <http://arxiv.org/abs/1803.06745>
- [120] B. R. Chakravarthi, V. Muralidaran, R. Priyadharshini, and J. P. McCrae, “Corpus creation for sentiment analysis in code-mixed Tamil-English text,” in *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*. Marseille, France: European Language Resources association, May 2020, pp. 202–210. [Online]. Available: <https://www.aclweb.org/anthology/2020.sltu-1.28>
- [121] A. Hande, R. Priyadharshini, and B. R. Chakravarthi, “KanCMD: Kannada CodeMixed dataset for sentiment analysis and offensive language detection,” in *Proceedings of the Third Workshop on Computational Modeling of People’s Opinions, Personality, and Emotion’s in Social Media*, M. Nissim, V. Patti, B. Plank, and E. Durmus, Eds. Barcelona, Spain (Online): Association for Computational Linguistics, Dec. 2020, pp. 54–63. [Online]. Available: <https://aclanthology.org/2020.peoples-1.6>
- [122] B. R. Chakravarthi, N. Jose, S. Suryawanshi, E. Sherly, and J. P. McCrae, “A sentiment analysis dataset for code-mixed Malayalam-English,” in *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*. Marseille, France: European Language Resources association, May 2020, pp. 177–184. [Online]. Available: <https://www.aclweb.org/anthology/2020.sltu-1.25>
- [123] R. Priyadharshini, B. R. Chakravarthi, S. Thavareesan, D. Chinnappa, D. Thenmozhi, and R. Ponnusamy, “Overview of the dravidiancodemix 2021 shared task on sentiment detection in tamil, malayalam, and kannada,” ser. FIRE ’21. New York, NY, USA: Association for Computing Machinery, 2022, p. 4–6. [Online]. Available: <https://doi.org/10.1145/3503162.3503177>

- [124] B. R. Chakravarthi, R. Priyadharshini, S. Thavareesan, D. Chinnappa, D. Thenmozhi, E. Sherly, J. P. McCrae, A. Hande, R. Ponnusamy, S. Banerjee, and C. Vasantharajan, “Findings of the sentiment analysis of dravidian languages in code-mixed text,” *CoRR*, 2021, vol. abs/2111.09811. [Online]. Available: <https://arxiv.org/abs/2111.09811>
- [125] T. Pires, E. Schlinger, and D. Garrette, “How multilingual is multilingual bert?” in *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, A. Korhonen, D. R. Traum, and L. Màrquez, Eds. Association for Computational Linguistics, 2019, pp. 4996–5001. [Online]. Available: <https://doi.org/10.18653/v1/p19-1493>
- [126] K. Shanmugavadivel, M. Subramanian, P. K. Kumaresan, B. R. Chakravarthi, B. Bharathi, S. C. Navaneethakrishnan, L. S. Kumar, T. Mandl, R. Ponnusamy, V. Palanikumar, and M. B. Jagadeeshan, “Overview of the shared task on sentiment analysis and homophobia detection of youtube comments in code-mixed dravidian languages,” in *Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation, Kolkata, India, December 9-13, 2022*, ser. CEUR Workshop Proceedings, K. Ghosh, T. Mandl, P. Majumder, and M. Mitra, Eds., vol. 3395. CEUR-WS.org, 2022, pp. 80–91. [Online]. Available: <https://ceur-ws.org/Vol-3395/T2-1.pdf>
- [127] S. Satapara, S. Masud, H. Madhu, M. A. Khan, M. S. Akhtar, T. Chakraborty, S. Modha, and T. Mandl, “Overview of the HASOC subtracks at FIRE 2023: Detection of hate spans and conversational hate-speech,” in *Proceedings of the 15th Annual Meeting of the Forum for Information Retrieval Evaluation, FIRE 2023, Panjim, India, December 15-18, 2023*, D. Ganguly, S. Majumdar, B. Mitra, P. Gupta, S. Gangopadhyay, and P. Majumder, Eds. ACM, 2023, pp. 10–12. [Online]. Available: <https://doi.org/10.1145/3632754.3633277>
- [128] N. Craswell and D. Hawking, “Overview of the trec-2002 web track,” in *Proceedings of TREC-2002*, November 2002. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/overview-of-the-trec-2002-web-track/>
- [129] C. J. Fox, “A stop list for general text,” *SIGIR Forum*, 1990, vol. 24, no. 1-2, pp. 19–35. [Online]. Available: <https://doi.org/10.1145/378881.378888>



# List of Publications

- **Published papers (Journal)**

- Supriya Chanda, Sukomal Pal. **The effect of stopword removal on Information Retrieval for Code-Mixed data obtained via social media**, *Springer Nature Computer Science*. 2023
- Supriya Chanda, Anshika Mishra, Sukomal Pal. **Sentiment Analysis of Code-Mixed Dravidian Languages leveraging pre-trained model and word-level language tag**, *Natural Language Processing (NLP)*. 2024

- **Published papers (Conference)**

- Supriya Chanda, Abhishek Dhaka, Sukomal Pal. **Towards Safer Online Spaces: Deep Learning for Hate Speech Detection in Code-Mixed Social Media Conversations**, *In 16th ACM Web Science Conference (Websci Companion '24)*, May 21–24, 2024, Stuttgart, Germany. ACM, New York, NY, USA 7 Pages. <https://doi.org/10.1145/3630744.3663610>
- Supriya Chanda, Sacchit Sheth, Sukomal Pal. **Coarse and Fine-Grained Conversational Hate Speech and Offensive Content Identification in Code-Mixed Languages using Fine-Tuned Multilingual Embedding**, *In Forum for Information Retrieval Evaluation (Working Notes) (FIRE)*, 2022, CEUR-WS.org
- Supriya Chanda, S. Ujjwal, Shayak Das, Sukomal Pal **Fine-tuning Pre-Trained Transformer based model for Hate Speech and Offensive Content Identification in English Indo-Aryan and Code-Mixed (English-Hindi) languages**, *In Forum for Information Retrieval Evaluation (Working Notes) (FIRE)*, pp. 446-458, 2021, CEUR-WS.org
- Supriya Chanda, Anshika Mishra, Sukomal Pal **Sentiment Analysis and Homophobia detection of Code-Mixed Dravidian Languages leveraging pre-trained model and word-level language tag**, *In Forum for Information Retrieval Evaluation (Working Notes) (FIRE)*, 2022, CEUR-WS.org

- **Published Book Chapter**

- Supriya Chanda, Sukomal Pal. **Hate content identification in code-mixed social media data**, *Text and Social Media Analytics for Fake News and Hate Speech Detection*, CRC Press, 2024.

- **Journal under revision**

- Supriya Chanda, Sukomal Pal. **Word level Language Identification from social media code-mixed data leveraging Transformer-based model**. *Expert Systems*
- Supriya Chanda, Sukomal Pal. **Code Mixed Information Retrieval: Building a dataset and phonetic algorithms-based approaches**. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*