

## REFERENCES

- [1] Y. Li, Z. A. Bandar, and D. Mclean, "An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources." *IEEE Transactions on Knowledge Data Engineering*, vol.[2] 15, no. 4, pp. 871–882, 2003.
- [2] JCS Kadupitiya, Surangika Ranathunga, and Gihan Dias. "Sinhala short sentence similarity calculation using corpus-based and knowledge-based similarity measures". In *Proceedings of the 6th Workshop on South and Southeast Asian Natural Language Processing (WSSLLP 2016)*, pp. 44–53, 2016.
- [3] Anutharsha Selvarasa, Nilasini Thirunavukkarasu, Niveathika Rajendran, Chinthoorie Yogalingam, Surangika Ranathunga, and Gihan Dias. "Short Tamil sentence similarity calculation using knowledge-based and corpus-based similarity measures." In *Engineering Research Conference (MERCCon), 2017 Moratuwa*, pp. 443-448. IEEE, 2017.
- [4] Li, Yuhua, Zuhair Bandar, David McLean, and James O'shea. "A Method for Measuring Sentence Similarity and its Application to Conversational Agents." In *FLAIRS Conference*, pp. 820-825. 2004.
- [5] Y. Li, D. McLean, Z. A. Bandar, J. D. O'Shea, and K. Crockett, "Sentence similarity based on semantic nets and corpus statistics." *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 8, pp. 1138–1150, 2006.
- [6] Zhao, J., Zhu, T. T., & Lan, M. Ecnu: "One stone two birds: Ensemble of heterogeneous measures for semantic relatedness and textual entailment". *Proceedings of the SemEval*, pp. 271-277, 2014.
- [7] He, Hua, Kevin Gimpel, and Jimmy Lin. "Multi-perspective sentence similarity modeling with convolutional neural networks." In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pp. 1576-1586. 2015.
- [8] Kiros, Ryan, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Antonio Torralba, Raquel Urtasun, and Sanja Fidler. "Skip-thought vectors." *arXiv preprint arXiv:1506.06726*. 2015.
- [9] Kai Sheng Tai, Richard Socher, and Christopher D. Manning. "Improved semantic representations from tree-structured long short-term memory networks". In *Association for Computational Linguistics (ACL)*. pp 1556-1566. 2015.
- [10] He, Hua, and Jimmy Lin. "Pairwise word interaction modeling with deep neural networks for semantic similarity measurement." In *Proceedings of the 2016 conference of the north American chapter of the Association for Computational Linguistics: human language technologies*, pp. 937-948. 2016.
- [11] Xie, Jun, Bo Chen, Xinglong Gu, Fengmei Liang, and Xinying Xu. "Self-attention-based BiLSTM model for short text fine-grained sentiment classification." *IEEE Access* vol. 7: pp. 180558-180570. 2019.
- [12] Jonas Mueller and Aditya Thyagarajan. "Siamese Recurrent Architectures for Learning Sentence Similarity". In *Proceedings of the 30th AAAI Conference on*

Artificial Intelligence, February 12--17, Phoenix, Arizona, USA. pp. 2786-2792. 2016.

[13] Chandrasekaran, Dhivya, and Vijay Mago. "Evolution of Semantic Similarity - A Survey." arXiv preprint arXiv:2004.13820. 2020.

[14] Gomaa, W. H. and Fahmy, A. A, "A Survey of Text Similarity Approaches". International Journal of Computer Applications 68(13), pp. 13-18, 2013.

[15] Allison, Lloyd, and Trevor I. Dix. "A bit-string longest-common-subsequence algorithm." Information Processing Letters 23, no. 5: pp. 305-310. 1986.

[16] Levenshtein, Vladimir I. "Binary codes capable of correcting deletions, insertions, and reversals." In Soviet physics doklady, vol. 10, no. 8, pp. 707-710. 1966.

[17] Hall, Patrick AV, and Geoff R. Dowling. "Approximate string matching." ACM computing surveys (CSUR) 12, no. 4: pp. 381-402. 1980.

[18] Ratcliff, John W., and David E. Metzener. "Pattern-matching-the gestalt approach." Dr Dobbs Journal, Issue 13, no. 7: pp. 46. 1988.

[19] Krause, Eugene F. "Taxicab geometry." The Mathematics Teacher 66, no. 8: pp. 695-706. 1973.

[20] Wegner, Peter. "A technique for counting ones in a binary computer." Communications of the ACM 3, no. 5: pp 322. 1960.

[21] Danielsson, Per-Erik. "Euclidean distance mapping." Computer Graphics and image processing 14, no. 3: pp. 227-248. 1980.

[22] Jaccard, Paul. "Étude comparative de la distribution florale dans une portion des Alpes et des Jura." Bull Soc Vaudoise Sci Nat 37: pp. 547-579. 1901.

[23] Dice, Lee R. "Measures of the amount of ecologic association between species." Ecology 26, no. 3: pp. 297-302. 1945.

[24] Burgess, Curt, Kay Livesay, and Kevin Lund. "Explorations in context space: Words, sentences, discourse." Discourse Processes 25, no. 2-3: pp. 211-257. 1998.

[25] Foltz, Peter W., Walter Kintsch, and Thomas K. Landauer. "The measurement of textual coherence with latent semantic analysis." Discourse processes 25, no. 2-3: pp. 285-307. 1998.

[26] Turney, Peter D. "Mining the web for synonyms: PMI-IR versus LSA on TOEFL." In European conference on machine learning, Springer, Berlin, Heidelberg, pp. 491-502. 2001.

[27] Mihalcea, Rada, Courtney Corley, and Carlo Strapparava. "Corpus-based and knowledge-based measures of text semantic similarity." In Proceedings of the American Association for Artificial Intelligence. (Boston, MA), vol. 6, no. 2006, pp. 775-780. 2006.

- [28] Miller, George A. "WordNet: a lexical database for English." *Communications of the ACM* 38, no. 11, pp. 39-41, 1995.
- [29] Rajendran, S., S. Arulmozhi, B. Kumara Shanmugam, S. Baskaran, and S. Thiagarajan. "Tamil wordnet." In *Proceedings of the first international global WordNet conference*. Mysore, vol. 152, pp. 271-274. 2002.
- [30] L. Arulmozhi, PT. Pattabhi R. K. Rao, Sobha, "A Hybrid POS Tagger for a Relatively Free Word Order Language." *Proceedings of the First National Symposium on Modeling and Shallow Parsing of Indian Languages*, pp. 79-85, 2006.
- [31] Welgama, V., Herath, D. L., Liyanage, C., Udalamatta, N., Weerasinghe, R., & Jayawardana, T. "Towards a Sinhala wordnet". In *Proceedings of the Conference on Human Language Technology for Development*, 2011
- [32] Wijesiri, Indeewari, Malaka Gallage, Buddhika Gunathilaka, Madhuranga Lakjeewa, Daya Wimalasuriya, Gihan Dias, Rohini Paranavithana, and Nisansa De Silva. "Building a wordnet for sinhala." In *Proceedings of the Seventh Global WordNet Conference*, pp. 100-108. 2014.
- [33] Weerasinghe, Ruvan, Dulip Herath, and Viraj Welgama. "Corpus-based Sinhala lexicon." In *Proceedings of the 7th Workshop on Asian Language Resources (ALR7)*, pp. 17-23. 2009.
- [34] Resnik, Philip. "Using information content to evaluate semantic similarity in a taxonomy." In *Proceedings of the 14th International Joint Conference on AI*. arXiv preprint [cmp-lg/9511007](https://arxiv.org/abs/1905.11007). 1995.
- [35] Lin, Dekang. "An information-theoretic definition of similarity." In *Proceedings of the International Conference on Machine Learning*, vol. 98, no. 1998, pp. 296-304. 1998.
- [36] Jiang, J and Conrath, D. "Semantic similarity based on corpus statistics and lexical taxonomy". In *Proceedings of the International Conference on Research in Computational Linguistics*. arXiv preprint [cmp-lg/9709008](https://arxiv.org/abs/1907.09008). 1997.
- [37] Leacock, C and Chodorow, M. "WordNet: An electronic lexical database". MIT Press, Chapter Combining local context and WordNet similarity for word sense identification, pp. 265–283. 1998.
- [38] Wu, Zhibiao, and Martha Palmer. "Verb semantics and lexical selection." In *Proceedings of the Annual Meeting Association for Computational Linguistics*. arXiv preprint [cmp-lg/9406033](https://arxiv.org/abs/1904.06033). 1994.
- [39] Lesk, Michael. "Automatic sense disambiguation using machine readable dictionaries: how to tell a pinecone from an ice cream cone." In *Proceedings of the 5th annual international conference on Systems documentation*, pp. 24-26. 1986.
- [40] Islam, Aminul, and Diana Inkpen. "Semantic text similarity using corpus-based word similarity and string similarity." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 2, no. 2: pp. 1-25. 2008.
- [41] Allison, Lloyd, and Trevor I. Dix. "A bit-string longest-common-subsequence algorithm." *Information Processing Letters* 23, no. 5: pp. 305-310. 1986.

- [42] Pawar, Atish, and Vijay Mago. "Calculating the similarity between words and sentences using a lexical database and corpus statistics." *IEEE Transactions on Knowledge and data Engineering*. arXiv preprint arXiv:1802.05667. 2018.
- [43] Lintean, Mihai C., and Vasile Rus. "Measuring Semantic Similarity in Short Texts through Greedy Pairing and Word Semantics." In *Flairs conference*, pp. 244-249. 2012.
- [44] Ferreira, Rafael, Rafael Dueire Lins, Steven J. Simske, Fred Freitas, and Marcelo Riss. "Assessing sentence similarity through lexical, syntactic and semantic analysis." *Computer Speech & Language* 39: pp. 1-28. 2016.
- [45] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* vol. 9, no. 8: pp. 1735-1780. 1997.
- [46] Graves, Alex, and Jürgen Schmidhuber. "Framewise phoneme classification with bidirectional LSTM and other neural network architectures." *Neural networks* vol. 18, no. 5-6 (2005): pp. 602-610. 2005.
- [47] Cho, Kyunghyun, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. "Learning phrase representations using RNN encoder-decoder for statistical machine translation." arXiv preprint arXiv:1406.1078. 2014.
- [48] Bromley, Jane, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. "Signature verification using a " siamese" time delay neural network." *Advances in neural information processing systems*: pp. 737-737. 1994.
- [49] Xie, Niantao, Sujian Li, and Jinglin Zhao. "ERCNN: Enhanced Recurrent Convolutional Neural Networks for Learning Sentence Similarity." In *18th China National Conference on Chinese Computational Linguistics*, pp. 119-130. 2019.
- [50] Yin, Wenpeng, Hinrich Schütze, Bing Xiang, and Bowen Zhou. "ABCNN: Attention-based convolutional neural network for modeling sentence pairs." *Transactions of the Association for Computational Linguistics* 4: pp. 259-272. 2016.
- [51] Liang, Zhiyao, and Jian Liu. "Sentence Similarity Measurement with Convolutional Neural Networks Using Semantic and Syntactic Features." *Computers, Materials & Continua*, vol.63, no.2: pp. 943-957. 2020.
- [52] Chen, Qin, Qinmin Hu, Jimmy Xiangji Huang, and Liang He. "CA-RNN: using context-aligned recurrent neural networks for modeling sentence similarity." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1. pp.9 2018.
- [53] Schuster, Mike, and Kuldip K. Paliwal. "Bidirectional recurrent neural networks." *IEEE transactions on Signal Processing* vol. 45, no. 11: pp. 2673-2681. 1997.
- [54] Chen, Qian, Xiaodan Zhu, Zhenhua Ling, Si Wei, Hui Jiang, and Diana Inkpen. "Enhanced lstm for natural language inference." arXiv:1609.06038. 2016.

- [55] Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781. 2013a.
- [56] Ichida, Alexandre Yukio, Felipe Meneguzzi, and Duncan D. Ruiz. "Measuring semantic similarity between sentences using a siamese neural network." In 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1-7. IEEE, 2018.
- [57] Ranasinghe, Tharindu, Constantin Orasan, and Ruslan Mitkov. "Semantic textual similarity with Siamese neural networks." In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019), pp. 1004-1011. 2019.
- [58] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. "Distributed representations of words and phrases and their compositionality". In Advances in neural information processing systems, pp. 3111-3119. 2013b.
- [59] Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. "Glove: Global vectors for word representation." In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532-1543. 2014.
- [60] T. Kenter and M. de Rijke. "Short text similarity with word embeddings". In CIKM, ACM, pp 1411-1420. 2015.
- [61] McCann, Bryan, James Bradbury, Caiming Xiong, and Richard Socher. "Learned in translation: Contextualized word vectors." In NIPS. arXiv preprint arXiv:1708.00107. 2017.
- [62] Mikolov, Tomáš, Wen-tau Yih, and Geoffrey Zweig. "Linguistic regularities in continuous space word representations." In Proceedings of the 2013 conference of the North American chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 746-751. 2013.
- [63] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. "Enriching word vectors with subword information". Transactions of the Association for Computational Linguistics 5 (2017), pp. 135-146. 2017.
- [64] Ranasinghe, Tharindu, Constantin Orasan, and Ruslan Mitkov. "Enhancing unsupervised sentence similarity methods with deep contextualised word representations." In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019), 2019.
- [65] Peters, Matthew E., Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. "Deep contextualized word representations." In NAACL. 2018a. arXiv preprint arXiv:1802.05365. 2018.
- [66] Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. "Language models are unsupervised multitask learners." OpenAI blog vol. 1, no. 8: pp. 9. 2019
- [67] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805. 2018.

- [68] Kevin Clark, Minh-Thang Luong, Christopher D Manning, and Quoc Le. "Semi-supervised sequence modeling with cross-view training". In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 1914 - 1925. 2018.
- [69] Howard, Jeremy, and Sebastian Ruder. "Universal language model fine-tuning for text classification." In ACL. Association for Computational Linguistics. arXiv preprint arXiv:1801.06146. 2018.
- [70] Marco, Marelli, Bentivogli Luisa, Baroni Marco, Bernardi Raffaella, Menini Stefano, and Zamparelli Roberto. "SemEval-2014 Task 1: Evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment." In Proc. SemEval, pp. 1-8. 2014.
- [71] Marelli, Marco, Stefano Menini, Marco Baroni, Luisa Bentivogli, Raffaella Bernardi, and Roberto Zamparelli. "A SICK cure for the evaluation of compositional distributional semantic models." In Lrec, pp. 216-223. 2014.
- [72] Van der Maaten, Laurens, and Geoffrey Hinton. "Visualizing data using t-SNE." Journal of machine learning research 9, no. 11, pp. 2579-2605, 2008.
- [73] Lakmal, Dimuthu, Surangika Ranathunga, Saman Peramuna, and Indu Herath. "Word embedding evaluation for sinhala." In Proceedings of the 12th Language Resources and Evaluation Conference, pp. 1874-1881. 2020.
- [74] "Tamil Stopwords list", TamilNLP, 2018. [Online]. Available: <https://github.com/AshokR/TamilNLP/wiki/Stopwords> [Accessed: Jul.26 2018].
- [75] Armand Joulin, Edouard Grave, Piotr Bojanowski, Tomas Mikolov, "Bag of Tricks for Efficient Text Classification". arXiv preprint 1607.01759, 2016.
- [76] Neculoiu, Paul, Maarten Versteegh, and Mihai Rotaru. "Learning text similarity with siamese recurrent networks." In Proceedings of the 1st Workshop on Representation Learning for NLP, pp. 148-157. 2016.