

LIST OF REFERENCE

- Afolabi, V., 2019, Pre-Processing Text For NLP, <http://www.vanaudelanalytix.com/python-blog/pre-processing-text-for-nlp>.
- Al-rimy, B., Maarof, M., Alazab, M., Alsolami, F., Mohd Shaid, S. Z., Ghaleb, F., Al-Hadhrani, T. & Ali, A., 2020, *A Pseudo Feedback-Based Annotated TF-IDF Technique for Dynamic Crypto-Ransomware Pre-Encryption Boundary Delineation and Features Extraction*, IEEE Access, 8(1), 1-1. 10.1109/ACCESS.2020.3012674.
- Ali, A. & Taqa, A., 2022, *Analytical Study of Traditional and Intelligent Textual Plagiarism Detection Approaches*, Journal Of Education And Science, 31(1), 8-25. 10.33899/edusj.2021.131895.1192.
- Ali, A., Alfayez, F. and Alquhayz, H., 2018. Semantic similarity measures between words: A brief survey. *Sci. Int.(Lahore)*, 30(6), pp.907-914.
- Anderson, T., 2021, Federalist Papers, <https://www.kaggle.com/datasets/tobyanderson/federalist-papers>.
- Anh, V.N. and Moffat, A., 2005, *Inverted index compression using word-aligned binary codes*, Information Retrieval, 8(1), pp.151-166.
- Asim A. M., El-Tahir, H. A., Dahwa M., and V'aclav S., 2017. Overview and Comparison of Plagiarism Detection Tools. VSB-Technical University of Ostrava.
- Blake, R., 2023, *Understanding Cosine Similarity and Its Applications*, <https://builtin.com/machine-learning/cosine-similarity>.
- Cutting D. and Pedersen J., 1990, *Optimization for Dynamic Inverted Index Maintenance*, Xerox Palo Alto Research Center.
- Devika, R., Vairavasundaram, S., Mahenthara, C.S.J., Varadarajan, V. and Kotecha, K., 2021. A deep learning model based on BERT and sentence transformer for semantic keyphrase extraction on big social data. *IEEE Access*, 9, pp.165252-165261.
- Divya, K.S., Subha, R. and Palaniswami, S., 2014. Similar words identification using naive and tf-idf method. *International Journal of Information Technology and Computer Science (IJITCS)*, 6(11), p.42.
- El-Rashidy, M.A., Mohamed, R.G., El-Fishawy, N.A. & Shouman, M.A., 2022, *Reliable plagiarism detection system based on deep learning approaches - Neural Computing and Applications*, Neural Computing and Applications 34, 18837-18858. <https://doi.org/10.1007/s00521-022-07486-w>.
- Eler, D.M., Grosa D., Pola I., Garcia R., Correia R., and Teixeira J., 2018, *Analysis of Document Pre-Processing Effects in Text and Opinion Mining*. Information, 9(4):100. <https://doi.org/10.3390/info9040100>.
- Fernando, S. & Stevenson, M., 2009, A Semantic Similarity Approach to Paraphrase Detection. *Proceedings of the 11th Annual Research Colloquium of the UK Special Interest Group for Computational Linguistics*.

- Gain A., Rao, M., and Bhat S., 2019, *Usage of Grammarly – Online Grammar and Spelling Checker Tool at the Health Sciences Library*, Manipal Academy of Higher Education, Manipal: A Study. *Library Philosophy and Practice (ejournal)* Vol. 4 No. 17, 2611, <https://digitalcommons.unl.edu/libphilprac/2611>.
- Gavilan C., 2015, *Information Ethics and Information Value*, Retrived November 28, 2023, from <https://www.gavilan.edu/library/LIB99/plagiarism.html>.
- Ghufron, A. M. & Rosyida, F., 2018, *The role of Grammarly in assessing English as a Foreign Language (EFL) writing*, *Lingua Cultura*, 12 (4).
- Goldstone, R.L., Medin, D.L. and Halberstadt, J., 1997. Similarity in context. *Memory & Cognition*, 25(2), pp.237-255.
- Gosa., 2022, *Information Storage and Retrieval*, Technology Institute, https://wachemo-learning.net/courses/information-storage-and-retrievalitec3081/#tab-course-section__overview.
- Gomede, E., 2023, *Exploring the Depths of Meaning: Semantic Similarity in Natural Language Processing*, <https://medium.com/@evertongomede/exploring-the-depths-of-meaning-semantic-similarity-in-natural-language-processing-19281e58558e>.
- Gunawan, D., Sembiring, C.A. and Budiman, M.A., 2018, *The Implementation Of Cosine Similarity To Calculate Text Relevance Between Two Documents*, *Journal Of Physics: Conference Series* (Vol. 978, No. 1, p. 012120). IOP Publishing.
- Hamilton, A., Madison, J., and Jay, J., 1788, *The Federalist*, The Independent Journal, United States.
- Henderi, H. and Winarno, W., 2021, *Text Mining an Automatic Short Answer Grading (ASAG), Comparison of Three Methods of Cosine Similarity, Jaccard Similarity and Dice's Coefficient*, *Journal of Applied Data Sciences*, 2(2).
- Katrin K., and Debora W., 2010, *Plagiarism Detection Test*, [https://plagiat.htw-berlin.de/wpcontent/uploads/Plagiarism DetectionTest2010-final.pdf](https://plagiat.htw-berlin.de/wpcontent/uploads/Plagiarism%20DetectionTest2010-final.pdf).
- Kong, Lei-lei, Han, Zhong-yuan., Qi, Hao-liang., Yang, Mu-yun., 2019, *Source Retrieval Model Focused on Aggregation for plagiarism detection*, *Information Sciences*, Volume 503.
- Krzyszewska U, Poniszewska-Marańda A, Ochelska-Mierzejewska J., 2022, *Systematic Comparison of Vectorization Methods in Classification Context*, *Applied Sciences*, 12(10):5119. <https://doi.org/10.3390/app12105119>.
- Lamba, Harshali & Govilkar, Sharvari. (2017). A Survey on Plagiarism Detection Techniques for Indian Regional Languages. *International Journal of Computer Applications*. 164. 44-50. 10.5120/ijca2017913622.
- Laskar, M.T.R., Huang, X. and Hoque, E., 2020, May. Contextualized embeddings based transformer encoder for sentence similarity modeling in answer selection task. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 5505-5514).
- Lebanoff, L., Derroncourt, F., Kim, D.S., Wang, L., Chang, W. and Liu, F., 2020. Learning to fuse sentences with transformers for summarization. *arXiv preprint arXiv:2010.03726*.

- Manning, C.D., Raghavan, P. and Schütze, H., Introduction to information retrieval? cambridge university press 2008. *Ch, 20*, pp.405-416.
- Nikolaev, D. and Padó, S., 2023. Representation biases in sentence transformers. *arXiv preprint arXiv:2301.13039*.
- Олізаренко, С., 2021. Method for determining the semantic similarity of arbitrary length texts using the transformers models. *Advanced Information Systems*, 5(2), pp.126-130.
- Qurashi, A.W., Holmes, V. and Johnson, A.P., 2020, August. Document processing: Methods for semantic text similarity analysis. In *2020 International Conference on Innovations in Intelligent Systems and Applications (INISTA)* (pp. 1-6). IEEE.
- S. Santini and R. Jain, "Similarity measures," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 871-883, Sept. 1999, doi: 10.1109/34.790428.
- Salton, G., Wong, A. and Yang, C.S., 1975. A Vector Space Model for Automatic Indexing "Communications of the ACM Vol. 18.
- Sanh, V., Debut, L., Chaumond, J. and Wolf T., 2019, DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.
- Schein M. and Paladugu R., 2001. Redundant surgical publications: tip of the iceberg? *Surgery Journal* Vol. 129 Issue 6, P655-611 2001, 10.1067/msy.2001.114549.
- Singh, A.K. and Shashi, M., 2019. Vectorization of text documents for identifying unifiable news articles. *International Journal of Advanced Computer Science and Applications*, 10(7). Singh, R. and Singh, S., 2021. Text similarity measures in news articles by vector space model using NLP. *Journal of The Institution of Engineers (India): Series B*, 102(2), pp.329-338.
- The University of Melbourne, 2016, *Verbatim copying/direct copying example*, Retrieved November 28, 2023, from <https://academicintegrity.unimelb.edu.au/plagiarism-and-collusion/verbatim-copying>.
- Thongtan, T. and Phienthrakul, T., 2019, July. Sentiment classification using document embeddings trained with cosine similarity. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop* (pp. 407-414).
- Tsang, S., 2022, *Review —TinyBERT: Distilling BERT for Natural Language Understanding*, <https://sh-tsang.medium.com/review-tinybert-distilling-bert-for-natural-language-understanding-6c49ad03fa94>.
- Wang, Z., Cui, J. and Zhu, Y., 2020, Plant recognition based on Jaccard distance and BOW, *Multimedia Systems* 26, 495–508, <https://doi.org/10.1007/s00530-020-00657-6>.
- Zouhir, Assia., Ayachi, R., Biniz, Mohamed., 2020, *A comparative Plagiarism Detection System methods between sentences*. Conference: The International Conference on Mathematics & Data Science (ICMDS) 2.