

IV. CONCLUSION

This study presents a systematic literature review to determine what cross language plagiarism detection techniques have been employed in the Natural Language context. We started with 130K potentially relevant studies, and after applying our inclusion/exclusion criteria and snowballing, we concluded our SLR with 107 documents. We provided an overview of different techniques to tackle cross-language plagiarism detection. Our results show that most of the selected studies follow a Machine Translation + Monolingual Analysis (T+MA). It means they first translated the suspicious and original documents into a common language (e.g., English) and then applied any monolingual techniques, e.g., fingerprints.

Many of the selected studies used “corpus” as the artifact in the evaluation phase. And we also observed there is no specific supporting tool; a few papers present some minimum details about its generic tool. Finally, we identified the implementation of any of these cross-language plagiarism detection techniques in a Software Engineering context as future work. Online communities like Stack Overflow present the problem of cross-language posts among SO websites.

ACKNOWLEDGMENT

This research was supported by the SENESCYT-Ecuador (scholar- ship program 2013-2).

REFERENCES

- [1] IEEE, “A Plagiarism FAQ,” 2015. [Online]. Available: http://www.ieee.org/publications_standards/publications/rights/plagiarism_FAQ.html. [Accessed: 11-May-2018].
- [2] M. Potthast, B. Stein, A. Barrón-Cedeño, and P. Rosso, “An evaluation framework for plagiarism detection,” in *Coling 2010: Posters*, 2010, pp. 997–1005.
- [3] S. M. Alzahrani, N. Salim, and A. Abraham, “Understanding plagiarism linguistic patterns, textual features, and detection methods,” *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 42, no. 2, pp. 133–149, 2012.
- [4] A. Barrón-Cedeño, P. Gupta, and P. Rosso, “Methods for cross-language plagiarism detection,” *Knowledge-Based Syst.*, vol. 50, pp. 211–217, 2013.
- [5] M. Franco-Salvador, P. Rosso, and M. Montes-y-Gómez, “A systematic study of knowledge graph analysis for cross-language plagiarism detection,” *Inf. Process. Manag.*, vol. 52, no. 4, pp. 550–570, 2016.
- [6] J. Ferrero, L. Besacier, D. Schwab, and F. Agnès, “Deep Investigation of Cross-Language Plagiarism Detection Methods,” in *Proceedings of the 10th Workshop on Building and Using Comparable Corpora*, 2017, pp. 6–15.
- [7] A. E. Tlitova, A. S. Toshev, M. Talanov, and V. Kurnosov, “Meta-Analysis of Cross-Language Plagiarism and Self-Plagiarism Detection Methods for Russian-English Language Pair,” *Front. Comput. Sci.*, vol. 2, p. 523053, 2020.
- [8] A. Kumar and S. Das, “An evolutionary survey from Monolingual Text Reuse to Cross Lingual Text Reuse in context to English-Hindi,” *Int. J. Sci. Eng. Res.*, vol. 6, no. 2, pp. 996–1003, 2015.
- [9] S. Shimpikar and S. Govilkar, “A Survey of Text Summarization Techniques for Indian Regional Languages,” *Int. J. Comput. Appl.*, vol. 165, no. 11, pp. 29–33, 2017.
- [10] P. Rosso, “Author profiling and Plagiarism detection,” in *Communications in Computer and Information Science*, 2015, vol. 505, pp. 229–250.
- [11] C. Wohlin, “Guidelines for snowballing in systematic literature studies and a replication in software engineering,” in *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, 2014, pp. 1–10.
- [12] J. Webster and R. T. Watson, “Analyzing the past to prepare for the future: Writing a literature review,” *MIS Q.*, pp. xiii–xxiii, 2002.
- [13] Keele University, “Guidelines for performing systematic literature reviews in software engineering,” 2007.
- [14] A. Martín-Martín, M. Thelwall, E. Orduna-Malea, and E. D. López-Cózar, “Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations’ COCI: a multidisciplinary comparison of coverage via citations,” *Scientometrics*, vol. 126, no. 1. Springer, pp. 907–908, 2021.
- [15] D. Landman, A. Serebrenik, and J. J. Vinju, “Challenges for static analysis of java reflection-literature review and empirical study,” in *2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE)*, 2017, pp. 507–518.
- [16] J. Cohen, “A coefficient of agreement for nominal scales,” *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960.
- [17] J. R. Landis and G. G. Koch, “The measurement of observer agreement for categorical data,” *Biometrics*, pp. 159–174, 1977.
- [18] L. Barbosa and J. Feng, “Robust sentiment detection on twitter from biased and noisy data,” in *Coling 2010: Posters*, 2010, pp. 36–44.
- [19] A. Barrón Cedeño, “On the Mono- and cross-language detection of text re-use and plagiarism,” *Proces. Leng. Nat.*, vol. 50, pp. 103–105, 2013.
- [20] M. Potthast, A. Barrón-Cedeño, B. Stein, and P. Rosso, “Cross-language plagiarism detection,” *Language Resources and Evaluation*, vol. 45, no. 1, pp. 45–62, 2011.
- [21] J. Kasprzak and M. Brandejs, “Improving the reliability of the plagiarism detection system: Lab report for PAN at CLEF 2010,” in *CEUR Workshop Proceedings*, 2010, vol. 1176, pp. 359–366.
- [22] C. K. Kent and N. Salim, “Web based cross language plagiarism detection,” in *2010 Second International Conference on Computational Intelligence, Modelling and Simulation*, 2010, pp. 199–204.
- [23] M. Pataki, “A new approach for searching translated plagiarism,” 2012.
- [24] C. K. Kent and N. Salim, “Web based cross language semantic plagiarism detection,” in *Proceedings - IEEE 9th International Conference on Dependable, Autonomic and Secure Computing, DASC 2011*, 2011, pp. 1096–1102.
- [25] S. Alzahrani, “Cross-Language Semantic Similarity of Arabic-English Short Phrases and Sentences,” *J. Comput. Sci.*, vol. 12, no. 1, pp. 1–18, 2016.
- [26] F. Safi-Esfahani, S. Sakian, and M. H. Nadimi-Shahraki, “English-Persian Plagiarism Detection based on a Semantic Approach,” *J. AI Data Min.*, vol. 5, no. 2, pp. 275–284, 2017.
- [27] R. Kothwal and V. Varma, “Cross lingual text reuse detection based on keyphrase extraction and similarity measures,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 7536 LNCS, pp. 71–78.
- [28] P. Gupta, K. Singhal, P. Majumder, and P. Rosso, “Detection of Paraphrastic Cases of Mono-lingual and Cross-lingual Plagiarism,” *ICON*, 2011.
- [29] P. Gupta and K. Singhal, “Mapping Hindi-English text re-use document pairs,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 7536 LNCS, pp. 79–85.
- [30] T. Brychcin, “Linear transformations for cross-lingual semantic textual similarity,” *Knowledge-Based Syst.*, vol. 187, p. 104819, 2020.
- [31] Z. Ceska, M. Toman, and K. Jezek, “Multilingual plagiarism detection,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008, vol. 5253 LNAI, pp. 83–92.
- [32] P. Gupta, A. Barrón-Cedeño, and P. Rosso, “Cross-language high similarity search using a conceptual thesaurus,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, vol. 7488 LNCS, pp. 67–75.
- [33] D. Pinto, J. Civera, A. Barrón-Cedeno, A. Juan, and P. Rosso, “A statistical approach to crosslingual natural language tasks,” *J. Algorithms*, vol. 64, no. 1, pp. 51–60, 2009.
- [34] S. Yahyaee, M. Bonzanini, and T. Roelleke, “Cross-lingual text fragment alignment using divergence from randomness,” in *International Symposium on String Processing and Information Retrieval*, 2011, pp. 14–25.
- [35] M. Mostafa and L. Agarwal, “Multilingual Plagiarism Detection,” 2014.
- [36] S. Alzahrani, N. Salim, A. A.-I. T. on, and undefined 2011, “Understanding plagiarism linguistic patterns, textual features and detection methods,” *researchgate.net*.

- [37] N. Ehsan, F. Tompa, ... A. S. the 2016 A. S. on, and undefined 2016, "Using a dictionary and n-gram alignment to improve fine-grained cross-language plagiarism detection," *dl.acm.org*.
- [38] P. Gupta, A. Barrón-Cedeno, and P. Rosso, "Cross-language high similarity search using a conceptual thesaurus," *Conf. Cross-Language*, 2012.
- [39] A. Barrón-Cedeno, P. Rosso, and E. Agirre, "Plagiarism detection across distant language pairs," *Proc. 23rd*, 2010.
- [40] J. Ferrero, L. Besacier, D. Schwab, and F. Agnes, "Deep Investigation of Cross-Language Plagiarism Detection Methods," in *Proceedings of the 10th Workshop on Building and Using Comparable Corpora*, 2017, pp. 6–15.
- [41] J. Ferrero, F. Agnès, L. Besacier, and D. Schwab, "A multilingual, multi-style and multi-granularity dataset for cross-language textual similarity detection," in *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*, 2016, pp. 4162–4169.
- [42] M. Pothast, A. Barrón-Cedeño, and B. Stein, "Cross-language plagiarism detection," *Lang. Resour.*, 2011.
- [43] L. T. Nguyen and D. Dien, "Vietnamese- English Cross-Lingual Paraphrase Identification Using Siamese Recurrent Architectures," in *Proceedings - 2019 19th International Symposium on Communications and Information Technologies, ISCIT 2019*, 2019, pp. 70–75.
- [44] J. Ferrero, F. Agnes, L. Besacier, and D. Schwab, "CompILIG at SemEval-2017 Task 1: Cross-language plagiarism detection methods for semantic textual similarity," *arXiv Prepr. arXiv1704.01346*, 2017.
- [45] E. M. B. Nagoudi, J. Ferrero, D. Schwab, and H. Cherroun, "Word embedding-based approaches for measuring semantic similarity of arabic-english sentences," in *International Conference on Arabic Language Processing*, 2017, pp. 19–33.
- [46] H. Ezzikouri, M. Erritali, and M. Oukessou, "Plagiarism Detection in Across Less Related Languages (English-Arabic): A Comparative Study," in *Smart Data and Computational Intelligence*, 2019, pp. 207–213.
- [47] C. Vania and M. Adriani, "Automatic external plagiarism detection using passage similarities," in *CEUR Workshop Proceedings*, 2010, vol. 1176.
- [48] E. Loginova, S. Varanasi, and G. Neumann, "Towards End-to-End Multilingual Question Answering," *Inf. Syst. Front.*, vol. 23, no. 1, pp. 227–241, 2021.
- [49] R. Billoshmi, R. Tripodi, and R. Navigli, "XL-AMR: Enabling cross-lingual AMR parsing with transfer learning techniques," in *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2020, pp. 2487–2500.
- [50] F. Issa, M. Damonte, S. B. Cohen, X. Yan, and Y. Chang, "Abstract meaning representation for paraphrase detection," in *NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 2018, vol. 1, pp. 442–452.
- [51] H. Asghari, O. Fatemi, S. Mohtaj, H. Faili, and P. Rosso, "On the use of word embedding for cross language plagiarism detection," *Intell. Data Anal.*, vol. 23, no. 3, pp. 661–680, 2019.
- [52] A. Micsik, P. Pallinger, and D. Siklósi, "Scaling a Plagiarism search service on the BonFIRE testbed," in *Proceedings of the International Conference on Cloud Computing Technology and Science, CloudCom*, 2013, vol. 2, pp. 57–62.
- [53] C. Chang, C.-H. Chang, and S.-Y. Hwang, "Employing word mover's distance for cross-lingual plagiarized text detection," *Proc. Assoc. Inf. Sci. Technol.*, vol. 57, no. 1, p. e229, 2020.
- [54] D. A. R. Torrejón, J. Manuel, and M. Ramos, "Detailed Comparison Module In CoReMo 1.9 Plagiarism Detector Notebook for PAN at CLEF 2012," *CLEF (Online Work. Notes/Labs/Workshop)*, pp. 1–8, 2012.
- [55] A. P. Zakiy Firdaus Alfikr, "The Construction of Indonesian-English Cross Language Plagiarism Detection System Using Fingerprinting Technique," *J. Comput. Sci. Inf.*, vol. 5, no. 1, pp. 16–23, 2012.
- [56] N. Ehsan and A. Shakery, "Candidate document retrieval for cross-lingual plagiarism detection using two-level proximity information," *Inf. Process. Manag.*, vol. 52, no. 6, pp. 1004–1017, 2016.
- [57] R. C. Pereira, V. P. Moreira, and R. Galante, "UFRGS @ PAN2010 : Detecting External Plagiarism," in *Lab Report for PAN at CLEF 2010*, 2010.
- [58] L. Gang, Z. Quan, and L. Guang, "Cross-language plagiarism detection based on WordNet," in *ACM International Conference Proceeding Series*, 2018, vol. Part F1376, pp. 163–168.
- [59] K. Mustofa and Y. A. Sir, "Early-Detection system for cross-language (translated) plagiarism," in *Information and Communication Technology-EurAsia Conference*, 2013, pp. 21–30.
- [60] L. T. Nguyen and D. Dien, "English-Vietnamese cross-language paraphrase identification method," in *ACM International Conference Proceeding Series*, 2017, vol. 2017-Decem, pp. 42–49.
- [61] A. Rücklé, N. S. Moosavi, and I. Gurevych, "Neural duplicate question detection without labeled training data," *arXiv Prepr. arXiv1911.05594*, 2019.
- [62] R. Lachraf, Y. Ayachi, A. Abdelali, D. Schwab, and others, "ArbEngVec: Arabic-English cross-lingual word embedding model," in *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, 2019, pp. 40–48.
- [63] A. Shojaei and F. Safi-Esfahani, "External Plagiarism Detection based on Human Behaviors in Producing Paraphrases of Sentences in English and Persian Languages," *J. AI Data Min.*, vol. 7, no. 3, pp. 451–466, 2019.
- [64] R. Bakhteev, A. Ogaltsov, A. Khazov, K. Safin, and R. Kuznetsova, "CrossLang: the system of cross-lingual plagiarism detection," *Work. Doc. Intell. NeurIPS 2019*, no. 18, pp. 1–5, 2019.
- [65] Z. Guan et al., "Cross-lingual multi-keyword rank search with semantic extension over encrypted data," *Inf. Sci. (Ny)*, vol. 514, pp. 523–540, 2020.
- [66] M. Franco-Salvador, P. Rosso, and R. Navigli, "A knowledge-based representation for cross-language document retrieval and categorization," in *14th Conference of the European Chapter of the Association for Computational Linguistics 2014, EACL 2014*, 2014, pp. 414–423.
- [67] A. A. Putri Ratna, F. Astha Ekadiyanto, I. Ibrahim, D. Husna, and F. Rahimullah, "Investigating Parallelization of Cross-language Plagiarism Detection System Using the Winnowing Algorithm in Cloud Based Implementation," in *2019 IEEE 10th International Conference on Awareness Science and Technology, iCAST 2019 - Proceedings*, 2019, pp. 1–7.
- [68] M. Pataki and A. C. Marosi, "Searching for Translated Plagiarism with the Help of Desktop Grids," *J. Grid Comput.*, vol. 11, no. 1, pp. 149–166, 2013.
- [69] J. Camacho-Collados, Y. Doval, E. Martínez-Cámara, L. Espinosa-Anke, F. Barbieri, and S. Schockaert, "Learning cross-lingual word embeddings from Twitter via distant supervision," in *Proceedings of the 14th International AAAI Conference on Web and Social Media, ICWSM 2020*, 2020, vol. 14, pp. 72–82.
- [70] S. Levy and W. Y. Wang, "Cross-lingual Transfer Learning for COVID-19 Outbreak Alignment," *arXiv Prepr. arXiv2006.03202*, 2020.
- [71] N. Poerner and H. Schütze, "Multi-view domain adapted sentence embeddings for low-resource unsupervised duplicate question detection," in *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 2020, pp. 1630–1641.
- [72] P. Rosso, "On the risk of cross-language plagiarism for less resourced languages such as Amazigh," *users.dsic.upv.es*, vol. 5, pp. 53–70, 2008.
- [73] I. Muneer, M. Sharjeel, M. Iqbal, R. M. A. Nawab, and P. Rayson, "CLEU - A Cross-language english-urdu corpus and benchmark for text reuse experiments," *J. Assoc. Inf. Sci. Technol.*, vol. 70, no. 7, pp. 729–741, 2019.
- [74] M. Pothast, B. Stein, and M. Anderka, "A Wikipedia-based multilingual retrieval model," in *European conference on information retrieval*, 2008, pp. 522–530.
- [75] D. Gupta, K. Vani, and C. K. Singh, "Using Natural Language Processing techniques and fuzzy-semantic similarity for automatic external plagiarism detection," in *Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014*, 2014, pp. 2694–2699.
- [76] D. V Zubarev and I. V Sochenkov, "Cross-language text alignment for plagiarism detection based on contextual and context-free models," in *Komp'juternaja Lingvistika i Intellektual'nye Tehnologii*, 2019, vol. 2019-May, no. 18, pp. 809–820.
- [77] N. Ehsan, A. Shakery, and F. W. Tompa, "Cross-lingual text alignment for fine-grained plagiarism detection," *J. Inf. Sci.*, vol. 45, no. 4, pp. 443–459, 2019.
- [78] M. Botto-Tobar, W. Torres, A. Lozano, M. G. J. van den Brand, B. Vasilescu, and A. Serebrenik, "Is stack overflow in portuguese attractive for brazilian users?," in *Proceedings of the 13th*

- International Conference on Global Software Engineering*, 2018, pp. 21–29.
- [79] B. Gipp, N. Meuschke, C. Breiting, J. Pitman, and A. Nürnberger, “Web-based demonstration of semantic similarity detection using citation pattern visualization for a cross language plagiarism case,” in *ICEIS 2014 - Proceedings of the 16th International Conference on Enterprise Information Systems*, 2014, vol. 2, pp. 677–683.
- [80] S. Alzahrani, N. Salim, C. K. Kent, M. S. Binwahlan, and L. Suanmali, “The development of cross-language plagiarism detection tool utilising fuzzy swarm-based summarisation,” in *Proceedings of the 2010 10th International Conference on Intelligent Systems Design and Applications, ISDA'10*, 2010, pp. 86–90.
- [81] H. Ezzikouri, M. Erritali, and M. Oukessou, “Plagiarism Detection in Across Less Related Languages (English-Arabic): A Comparative Study,” in *International Conference on Advanced Information Technology, Services and Systems*, 2018, pp. 207–213.
- [82] S. Parida, E. Villatoro-Tello, S. Kumar, P. Motlicek, and Q. Zhan, “Idiap Submission to Swiss-German Language Detection Shared Task,” in *SwissText/KONVENS*, 2020.
- [83] M. Roostae, M. H. Sadreddini, and S. M. Fakhrahmad, “An effective approach to candidate retrieval for cross-language plagiarism detection: A fusion of conceptual and keyword-based schemes,” *Inf. Process. & Manag.*, vol. 57, no. 2, p. 102150, 2020.
- [84] A. Shojaie and F. Safi-Esfahani, “External Plagiarism Detection based on Human Behaviors in Producing Paraphrases of Sentences in English and Persian Languages,” *J. AI Data Min.*, vol. 7, no. 3, pp. 451–466, 2019.
- [85] F. Ture, T. Elsayed, and J. Lin, “No free lunch: brute force vs. locality-sensitive hashing for cross-lingual pairwise similarity,” in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, 2011, pp. 943–952.
- [86] R. Pereira, V. Moreira, and R. Galante, “A new approach for cross-language plagiarism analysis,” *Conf. Cross-Language*, 2010.
- [87] B. Pouliquen, R. Steinberger, and C. Ignat, “Automatic identification of document translations in large multilingual document collections,” *arXiv Prepr. cs/0609060*, 2006.
- [88] Z. Ceska, M. Toman, and K. Jezek, “Multilingual plagiarism detection,” *Int. Conf. Artif.*, 2008.
- [89] V. Thompson, “Detecting cross-lingual plagiarism using simulated word embeddings,” *arXiv Prepr. arXiv1712.10190*, 2017.
- [90] J. Ray Chowdhury, C. Caragea, and D. Caragea, “Cross-lingual disaster-related multi-label tweet classification with manifold mixup,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, 2020.
- [91] S. Alzahrani and H. Aljuaid, “Identifying cross-lingual plagiarism using rich semantic features and deep neural networks: A study on Arabic-English plagiarism cases,” *J. King Saud Univ. Inf. Sci.*, 2020.
- [92] C. Lo and M. Simard, “Fully unsupervised crosslingual semantic textual similarity metric based on BERT for identifying parallel data,” in *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, 2019, pp. 206–215.
- [93] A. K. Khakimova, M. M. Charmine, A. A. Klovov, and E. G. Sokolov, “Approaches to assessing the semantic similarity of texts in a multilingual space,” 2020.
- [94] N. Alotaibi and M. Joy, “Using Sentence Embedding for Cross-Language Plagiarism Detection,” in *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, 2020, pp. 373–379.
- [95] M. Ustaszewski, “Exploring Adequacy Errors in Neural Machine Translation with the Help of Cross-Language Aligned Word Embeddings,” in *Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT 2019)*, 2019, pp. 122–128.
- [96] J. Ferrero, F. Agnes, L. Besacier, and D. Schwab, “Using word embedding for cross-language plagiarism detection,” *arXiv Prepr. arXiv1702.03082*, 2017.
- [97] A. A. P. Ratna *et al.*, “Cross-language plagiarism detection system using latent semantic analysis and learning vector quantization,” *Algorithms*, vol. 10, no. 2, p. 69, 2017.
- [98] A. A. P. Ratna *et al.*, “Cross-Language Automatic Plagiarism Detector Using Latent Semantic Analysis and Self-Organizing Map,” in *Proceedings of the 2018 International Conference on Artificial Intelligence and Virtual Reality*, 2018, pp. 83–87.
- [99] S. Srivastava and S. Govilkar, “A Survey on Paraphrase Detection Techniques for Indian Regional Languages,” *Int. J. Comput. Appl.*, vol. 163, no. 9, pp. 975–8887, 2017.
- [100] M. S. Arefin, Y. Morimoto, and M. A. Sharif, “BAENPD: A Bilingual Plagiarism Detector,” *J. Comput.*, vol. 8, no. 5, pp. 1145–1156, 2013.
- [101] M. Muhr and R. Kern, “External and intrinsic plagiarism detection using a cross-lingual retrieval and segmentation system,” in *2nd International Competition on Plagiarism Detection*, 2010.
- [102] Y. Qin, “Cross-Lingual Similarity Discrimination with Translation Characteristics,” *Int. J. Artif. Intell. & Appl.*, vol. 4, no. 5, p. 39, 2013.
- [103] A. Barrón-Cedeno, P. Rosso, D. Pinto, and A. Juan, “On Cross-lingual Plagiarism Analysis using a Statistical Model,” *PAN*, vol. 212, pp. 1–10, 2008.
- [104] M. Franco-Salvador, P. Gupta, and P. Rosso, “Knowledge graphs as context models: Improving the detection of cross-language plagiarism with paraphrasing,” in *PROMISE Winter School*, 2013, pp. 227–236.
- [105] M. Franco-Salvador, P. Gupta, ... P. R.-K.-B., and undefined 2016, “Cross-language plagiarism detection over continuous-space-and knowledge graph-based representations of language,” *Elsevier*.
- [106] N. Radoev, A. Zouq, and M. Gagnon, “Multilingual Question Answering using Lexico-Syntactic Patterns,” *Resource*, vol. 65, pp. 86–88.
- [107] M. Potthast, B. Stein, and M. Anderka, “A Wikipedia-based multilingual retrieval model,” *Eur. Conf. Inf.*, 2008.
- [108] H. Ezzikouri, M. Erritali, and M. Oukessou, “Fuzzy-semantic similarity for automatic multilingual plagiarism detection,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 9, pp. 86–90, 2017.
- [109] H. Ezzikouri, M. Oukessou, M. Youness, and M. Erritali, “Fuzzy cross language plagiarism detection (Arabic-English) using WordNet in a big data environment,” in *Proceedings of the 2018 2nd International Conference on Cloud and Big Data Computing*, 2018, pp. 22–27.
- [110] D. Dinh and N. Le Thanh, “English--Vietnamese cross-language paraphrase identification using hybrid feature classes,” *J. Heuristics*, pp. 1–17, 2019.
- [111] A. Barrón-Cedeno, P. Rosso, S. Devi, and P. Clough, “Pan@ fire: Overview of the cross-language! ndian text re-use detection competition,” *Inf. Access*, 2013.
- [112] J. Kasprzak and M. Brandejs, “Improving the Reliability of the Plagiarism Detection System,” in *Proceedings of the International Conference of the Cross-Language Evaluation Forum (CLEF 2010), Uncovering Plagiarism, Authorship, and Social Software Misuse Workshop (PAN'10)*, 2010, pp. 359–366.
- [113] R. Kothwal and V. Varma, “Cross lingual text reuse detection based on keyphrase extraction and similarity measures,” *Multiling. Inf. Access South Asian*, 2013.
- [114] D. A. R. Torrejón and J. M. M. Ramos, “Text alignment module in CoReMo 2.1 plagiarism detector,” *Forner et al. [34]*, 2013.
- [115] Z. Alaa, S. Tiun, and M. Abdulameer, “Cross-Language Plagiarism of Arabic-English Documents Using Linear Logistic Regression,” *J. Theor. & Appl. Inf. Technol.*, vol. 83, no. 1, 2016.
- [116] A. Aljohani and M. Mohd, “Arabic-English cross-language plagiarism detection using winnowing algorithm,” *Inf. Technol. J.*, vol. 13, no. 14, p. 2349, 2014.
- [117] M. Al-suhaiqi11, M. A. S. Hazaa22, and M. Albared33, “Arabic English Cross-Lingual Plagiarism Detection Based on Keyphrases Extraction, 2 Monolingual and Machine Learning Approach 3,” 2018.
- [118] M. Sharjeel, *Mono-and cross-lingual paraphrased text reuse and extrinsic plagiarism detection*. Lancaster University (United Kingdom), 2020.
- [119] A. Rücklé, K. Swarnkar, and I. Gurevych, “Improved cross-lingual question retrieval for community question answering,” in *The world wide web conference*, 2019, pp. 3179–3186.
- [120] R. Jungnickel, A. Pomp, A. Kirmse, X. Li, V. Samsonov, and T. Meisen, “Evaluation and Comparison of Cross-lingual Text Processing Pipelines,” in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2019, pp. 417–425.
- [121] L. Gang, Z. Quan, and Y. Qianru, “Cross-language plagiarism detection technology based on fingerprint fusion”.