Abstract

In this article we will discuss the work of a semantic comparison method, through which the detection of plagiarism is revealed in the fuzzy information, which we have designed an algorithm with a semantic dimension to detect plagiarism in the fuzzy information and detect impersonations such as changing the structure of speech or replacing words with synonyms and limiting technical spelling errors such as not completely writing the end of the word or unofficial and unknown abbreviations, and analyze shows the degree of similarity of the original text, and analyzes the overall evaluation of the degree of similarity of texts from the apparent structures of the text. Experiments have shown that the proposed method with a semantic dimension in the case of fuzzy information is better than sherlock method in terms of file size criterion of 6% if using word synonyms in the file and 1% in case of rewriting the file. As for the standard time taken to examine the files through the acceleration calculation, it is noted that the proposed method for the semantic dimension in the case of fuzzy information is faster in performance than the Sherlock method in the case of the use of synonyms 1.02 times and in
Designing Method and Algorithm of Semantic Comparison for Detecting Similarities in the Case of Fuzzy Information using Software Engineering and Medical Ontologies

the case of rewriting words with a value of 1.01 times in the case of file size 382 words. The results of the experiments show that the average execution time of the proposed algorithm, for finding plagiarism, is less by 3.47% compared with the Sherlock algorithm in the case of the use of synonyms and less by 1.83% compared with the Sherlock algorithm in the case of rewriting words. The algorithm works effectively as the file size increases, the gain ratio is obtained up to 2.73% in the synonym of words and 2.69% in the case of rewriting words. From the results presented in the tables, we conclude that the average error rate of the proposed algorithm is 2% lower than the error rate sherlock algorithm. The complexity of the proposed algorithm is O(m*n).

References

7. Bela Gipp. 2014. Citation-based Plagiarism Detection - Detecting Disguised and Cross-language Plagiarism using Citation Pattern Analysis. Springer.
S. "Intelligent plagiarism detection mechanism using semantic technology: A different approach". International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2013 Aug 22 (pp. 779-783). IEEE.


15. RDF/OWL Representation of WordNet. [online] Available at: http://www.w3.org/TR/wordnet-rdf/ [Accessed 1-may 2015].


30. The Open Biological and Biomedical Ontologies. [online] Available at: http://www.obofoundry.org/cgi-bin/detail.cgi?id=OGMS [Accessed 25-may 2015].


Index Terms

Computer Science  Algorithms

Keywords

Plagiarism detection, medical ontologies, software engineering ontology, Semantic network.