

## **Plagiarism Detection in Arabic Language using Rhetorical Structure Theory\***

**Eng. Khaled Omar\*\***

**Dr. Bassel Alkhatib\*\*\***

---

### **Abstract**

This paper presents a review of available algorithms and plagiarism detection systems, and an implementation of Plagiarism Detection System using available search engines on the web.

Plagiarism detection in natural language documents is a complicated problem and it is related to the characteristics of the language itself.

There are many available algorithms for plagiarism detection in natural languages .Generally these algorithms belong to two main categories ; the first one is plagiarism detection algorithms based on fingerprint and the second is plagiarism detection algorithms based on content comparison and includes string matching and tree matching algorithms .

Usually available systems of plagiarism detection use specific type of detection algorithms or use a mixture of detection algorithms to achieve effective detection systems (fast and accurate).

In this research, a plagiarism detection system has been developed using Bing search engine and a plagiarism detection algorithm based on Rhetorical Structure Theory.

---

**Keywords:** Arabic plagiarism detection, Rhetorical Structure Theory, Natural language processing.

---

\* For The paper in Arabic see pages (73-83)

\*\*Damascus university-Informatics Engineering College

\*\*\*Damascus university-Informatics Engineering College

## References

1. Vinod K.R., Sandhya.S, Sathish Kumar D, Harani A, David Banji, Otilia JF Banji, Plagiarism-history detection and prevention, Journal for drugs and medicines, Vol.3, Issue:1, pp.1- 4, 2011.
2. Shizhong Wu; Yongle Hao; Xinyu Gao; Baojiang Cui; Ce Bian, Homology Detection Based on Abstract Syntax Tree Combined Simple Semantics Analysis, Web Intelligence and Intelligent Agent Technology (WI-IAT), vol.3, pp.410-414, 2010
3. Al-Khatib B., Aspel A., Saleh M., fares M., Hamad M.M., plagiarism detection using the web, Damascus university,informatics engineering college, 2007.
4. Al-Sanie W., Towards an infrastructure for Arabic text Summarization using Rhetorical Structure Theory, master thesis , king Saud University, K.S.A., 2005
5. Bing , API Basics. [online] Available at: <http://www.bing.com/developers/s/APIBasics.html> [Accessed 15-October 2011 ].
6. Mathkour H., Touir A., Al-Sanea W., Parsing Arabic Texts Using Rhetorical Structure Theory, Journal of Computer Science , vol. 4, no. 9, pp. 713-720, 2008
7. Taboada M., Mann C. W. Rhetorical Structure Theory: looking back and moving ahead, journal of Discourse Studies, vol. (8),pp. 423—459, 2006.
8. Kent K. C., Salim N., Web Based Cross Language Plagiarism Detection, Journal of computing, vol.1, issue 1, issn: 2151-9617, 2009.
9. Mozgovoy M., Tusov V., Klyuev V., The Use of Machine Semantic Analysis in Plagiarism Detection, Proceedings of the 9th International Conference on Humans and Computers, Aizu-Wakamatsu, Japan, pp. 72-77, 2006.
10. Liu Y. , Liang L., A Dual-method Model for Copy Detection, Web Intelligence and Intelligent Agent Technology Workshops of IEEE/WIC/ACM International Conference, pp.634 – 637, 2006.
11. Schleimer S., Winnowing: Local Algorithms for Document Fingerprinting, Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, pp.76-85, 2003.
12. Touir A., Mathkour H., Al-Sanea W., Semantic-Based Segmentation of Arabic Texts, Information Technology Journal, Vol. 7, pp. 1009-1015, 2008.
13. Al-Sanea W., Mathkour H., Touir A, On an Arabic Text Segmentation Techniques, In proceeding of The Ninth International Conference on Information Integration and Web-based Applications Services, Jakarta, 2007.
14. Mathkour H., Member, IAENG, A Novel Rhetorical Structure Approach for Classifying Arabic Security Documents, International Journal of Computer Theory and Engineering, Vol. 1, No. 3,pp.195-200, 2009.
15. Iraky K., A Comprehensive Taxonomy of Arabic Discourse Coherence Relations, in Taibah University International Conference on Computing and Information Technology Al-Madinah Al-Munawwarah, Saudi Arabia, pp.491-496, 2012.
16. Mozgovoy M., Tuomo K., Erkki S., Using natural language parsers in plagiarism detection, in Speech and Language Technology in Education, pp.77-79, 2007.
17. Tachaphetpiboon S., Facundes N., Amornraksa T., Plagiarism Indication by Syntactic-Semantic Analysis, In proceeding of Communications, APCC 2007. Asia-Pacific Conference, 2007.