Detection of plagiarism in Arabic texts using text mining: a software agent based approach

Dr Abdullah Al Hussein - Assistant Professor
College of Computer and Information Sciences
Majmaah University, Saudi Arabia

Dr R. Jayabrabu - Associate Professor
Department of Computer Applications
Dr. N.G.P Institute of Technology, India

Dr Saravanan Venkataraman Tirumalai - Associate Professor
College of Computer and Information Sciences
Majmaah University, Saudi Arabia

Many tools are available to detect plagiarism in English. Detecting plagiarism in Arabic documents is particularly a challenging task because of the complex linguistic structure of Arabic. The objective is to develop a text mining based plagiarism detection tool with the help of software agents for processing Arabic text documents. Textmining in Rapidminer extracts the tokens from text-based content such as word documents and emails. The text mining thus reduces the time taken in the initial processing. Also, the token generated by the text mining tools are more clear and valid. An open source tool Rapidminer is used for text mining. As this tool, does not have a feature to process Arabic documents, a new add-on will be developed and attached with Rapidminer. The tool then processes the Arabic documents and stores the resultant tokens in the archive. The winnowing n-gram fingerprinting algorithm is analyzed, modified and applied to the archive to find the similarity index. This paper present plagiarism detection architecture for comparison of Arabic texts to identify similarities. The tool is based on textmining which reduces the time in preliminary part in the detection of plagiarisms. Software agents are also used for better comparisons and to find out more similar texts. Different types of agents such as User-interface agent, Textmining agent, Archiva-agent are used and procedures are developed for the implementation. The performance of this tool is evaluated on a large data set of Arabic texts. This
paper also presents the advantages of using the software agents with Textmining.