PhD Examination Report

Name: Hussein Khaled Hussein Soori

University: VSB Technical University of Ostrava

Thesis title: Data Compression Approach for Plagiarism Detection

Similarity detection is considered a crucial part of document processing. It covers a wide area including spam detection, plagiarism detection, etc. The need for plagiarism detection tools is growing with the growing number of natural language documents that are written day by day in the academic institutions all over the world. This thesis is devoted to create plagiarism detection tools by implementing a similarity measurement based on the initialization of Lempel Ziv comparison algorithms and its modifications and show its efficiency for detecting plagiarism of Arabic and Czech texts.

The proposed framework is used for Arabic language. Unlike languages that use Roman characters, Arabic is written from right to left and has twenty eight alphabets (three vowels and twenty five consonants). In case of Arabic, additional steps have to be performed before the computational processing such as, text normalization and some other steps that are specific to Arabic text processing. Hence, Arabic plagiarism detection tools require considering language-specific features in detecting text similarity.

The thesis is divided into four chapters. Chapter one is an introduction to the thesis and includes the objectives and the organization of the thesis. Chapter two deals with the syllabification method for Arabic language based on vowel letters and includes a light stemming method for data compression enhanced by an online parser for the categorization of Arabic parts of speech. Chapter three is divided into three parts. The first and second parts are dedicated to plagiarism detection tools for Arabic and Czech texts by implementing a similarity measurement based on a comparison algorithms. The third part includes a semantic plagiarism detection tool for reworded and restructured Czech texts by using a synonyms thesaurus and a stemmer. The last chapter includes the conclusions drawn from this thesis.
According to Scopus, the author has published a list of 8 publications, including one article in a Scopus journal. I think that the number of publications and their quality constitute a good standard for Ph.D. student.

I am convinced that the presented Ph.D. thesis represents good study providing valuable contribution to the state of scientific knowledge in the area. The author of the thesis proved his ability to conduct research and achieve scientific results. I am impressed by the quality of the work, which is indeed very tough. I believe that this thesis has made a good scientific contribution for the award of a PhD degree and I take this opportunity to congratulate the student and his supervisor.

Prof. (Dr.) Ajith Abraham
Director - Machine Intelligence Research Labs (MIR Labs)
Scientific Network for Innovation and Research Excellence
P.O. Box 2259
Auburn, Washington 98071, USA
http://www.mirlabs.org
Email: ajith.abraham@mirlabs.org / ajith.abraham@ieee.org
Tel: 001-253-2453001-- Fax: 001-615-694-4607

Personal WWW: http://www.softcomputing.net
Profile: http://www.researcherid.com/rid/A-1416-2008

September 02, 2016