

Abstract of thesis presented to the Senate of Universiti Malaysia Terengganu in fulfillment of the requirements for degree of Doctor of Philosophy in Computer Science

**A SEMANTIC REPRESENTATION FOR IMPROVING THE PERFORMANCE
OF COMPARATIVE QUESTION ANSWERING SYSTEMS**

IBRAHIM MAHMOUD IBRAHIM ALTURANI

JUNE 2021

Main Supervisor : Associate Prof. Mohd Pouzi bin Hamzah, Ph.D
Faculty : Faculty of Ocean Engineering Technology and Informatics

Question Answering (QA) is considered a vital research area in Information Retrieval (IR) and Artificial Intelligence (AI). Most QA models depend on bag-of-words as a representation of the documents and questions when answering more than just the semantic representation in relevance estimation of information which in turn reduces the effectiveness and performance of the QA systems. The aim of improving QA systems is to increase accurate answers and provide access to a large number of related passages. Overall, the QA system is incapable of understanding the natural language with high accuracy. This could be attributed to its richness of hidden and embedded information. Accordingly, the main aim of this research is to design a model that uses the semantic approach to search and retrieve relevant answers in the QA system, particularly for comparative questions. This research proposes a new model which will improve question classification and document representations by using semantic analyses. The classification of these questions can be improved with an additional classification layer that can be added in the question processing stage for further semantic classification. This layer will be responsible for grouping the question as comparative or non-comparative. The representation of the documents can be improved by utilizing useful

indexes of semantic relatedness and concepts in addition to keywords and lexical terms. This approach involves ontology development based on comparative classification. In this research, a sample is randomly selected from Smith dataset comprising 150 documents and 333-questions to evaluate the proposed model. The graph model and semantic similarity are adapted in indexing and extracting processes. The ultimate goal of the proposed model is to increase the ability of the graph model by improving the representation of different relationships among the concepts supported by WordNet. The proposed model demonstrated a significant improvement in the performance of QA system based on the semantic question classification and semantic representation of the relationships for the contents of the text documents by analyzing the relationships among different concepts to extract the answer. Recall and precision are used to investigate and evaluate the effectiveness of the proposed model. The results of the experiments have shown a significant improvement of the precision of the retrieval system using a semantic representation of up to 14% compared to the keyword-based searching method, 7.5% compared to query-expansion method and 2% compared to the semantic graph model. Regarding the recall, the proposed model achieved superior performance up to 16.8%, 8.7% and 1.5% over other methods respectively.

Abstrak tesis yang dikemukakan kepada Senat Universiti Malaysia
Terengganu sebagai memenuhi keperluan untuk ijazah Doktor Falsafah Sains Komputer

**PERWAKILAN SEMANTIK UNTUK MENINGKATKAN
PRESTASI SISTEM JAWAPAN PERTANYAAN PERBANDINGAN**

IBRAHIM MAHMOUD IBRAHIM ALTURANI

Jun 2021

Penyelia Utama : Prof Madya. Mohd Pouzi bin Hamzah, Ph.D
**Fakulti : Fakulti Teknologi Kejuruteraan Kelautan dan
Informatik**

Jawapan Pertanyaan (JP) merupakan bidang penyelidikan penting dalam capaian maklumat dan kepintaran buatan. Kebanyakan model JP bergantung kepada beg-perkataan sebagai perwakilan dokumen dan soalan bagi perwakilan semantik penganggaran maklumat relevan dan kaedah ini mengurangkan keberkesanan dan prestasi sistem JP. Penambahbaikan sistem JP adalah dengan meningkatkan ketepatan jawapan dan juga meningkatkan capaian kepada jawapan yang berkaitan dengan lebih luas. Secara umumnya sistem JP tidak dapat memahami bahasa tabii dengan kejituan yang tinggi. Ini disebabkan oleh maklumat yang terkandung biasanya dalam bentuk tidak langsung. Matlamat kajian ini adalah untuk merekabentuk model yang menggunakan pendekatan semantik untuk mencari jawapan relevan dalam sistem JP khususnya bagi soalan berbentuk perbandingan. Kajian ini mencadangkan model baru yang membolehkan peningkatan klasifikasi soalan dan perwakilan dokumen melalui analisis semantik. Klasifikasi soalan boleh diperbaiki dengan tambahan lapisan klasifikasi yang digunakan semasa fasa pemprosesan soalan bagi meningkatkan lagi klasifikasi semantik. Lapisan ini berfungsi sebagai pengkelas soalan perbandingan atau bukan perbandingan. Perwakilan dokumen boleh ditambahbaik dengan penggunaan

indeks semantik hubungan antara konsep sebagai tambahan kepada kata kunci dan perkataan leksikal. Pendekatan ini melibatkan pembangunan ontologi berasaskan klasifikasi perbandingan. Dalam kajian ini, sampel diambil secara rawak dari set data Smith yang terdiri dari 150 dokumen dan 333 soalan untuk menilai model yang dicadangkan. Model graf dan kesamaan semantik digunakan dalam pengindeksan dan proses pengekstrakan. Model yang dicadangkan ini meningkatkan keupayaan model graf dengan memperbaiki pelbagai perwakilan hubungan antara konsep yang disokong oleh WordNet. Model ini telah menunjukkan keberkesanan yang tinggi dalam sistem JP yang berasaskan klasifikasi semantik soalan dan perwakilan semantik hubungan antara kandungan dokumen melalui analisis hubungan antara konsep bagi mendapatkan jawapan. Pengukuran dapatan semula dan kejituan digunakan untuk menilai ketepatan model yang dicadangkan. Hasil eksperimen menunjukkan peningkatan ketara dalam kejituan menggunakan model yang dicadangkan sehingga 14% berbanding dengan kaedah carian menggunakan katakunci, 7.5% berbanding dengan kaedah pengembangan kueri dan 2% berbanding dengan model graf semantik. Bagi dapatan semula, model yang dicadangkan mencapai prestasi 16%, 8.7% dan 1.5% melebihi kaedah lain diatas.