

**META SEARCH ENGINE  
POWERED BY DBPEDIA**

**GRANT CODE: FRG0208-TK-1/2010**

**PROJECT LEADER:  
DR. CHIN KIM ON**

**CO-INVESTIGATOR:  
ASSOC. PROF. DR. PATRICIA ANTHONY**

**RESEARCH OFFICER:  
MR. BOO VOOI KEONG**

**FINAL RESEARCH REPORT**

**SCHOOL OF ENGINEERING AND  
INFORMATION TECHNOLOGY**

**UNIVERSITI MALAYSIA SABAH**

**2012**



**UMS**  
UNIVERSITI MALAYSIA SABAH

## SINOPSIS

Kesusahan untuk mencari maklumat dari laman web adalah setimpal dengan saiz maklumat yang terdapat dalam laman web. Pada masa sekarang, saiz laman web sudah mencapai suatu kadar di mana algoritma penilaian yang baik tidak mencukupi untuk menghasilkan hasil carian yang tepat. Kemungkinan besar, suatu perkataan yang boleh mewakili lebih daripada satu makna akan memberikan cabaran terhadap mesin carian untuk memaparkan hasil carian kerana mesin carian tradisi hanya menilai kunci perkataan pengguna berdasarkan perkataan dalam laman web, tanpa menilai maksud yang diminta oleh pengguna. Mesin carian semantik yang berdasarkan konsep untuk memahami maksud kunci perkataan pengguna pula berpotensi untuk menyelesaikan masalah tersebut. Walaubagaimanapun, laman web sekarang masih berada pada peringkat Web 2.0, di mana maklumat dan data dipaparkan dalam bentuk yang tidak berstruktur. Bentuk maklumat tersebut hanya sesuai untuk pemahaman manusia sahaja. Dengan ini, pembinaan suatu mesin carian yang semantik adalah suatu tugas yang mencabar. Untuk tujuan mencapai hasil carian berdasarkan matlamat carian semantik, terdapat banyak peningkatan perlu dicapai terutamanya dalam bidang pemprosesan bahasa. Dengan ini, penyelidikan ini mencadangkan suatu mesin carian meta yang dapat meningkatkan prestasi mesin carian tradisi dan memperkayakan hasil carian dengan mempertimbangkan maksud daripada kunci perkataan pengguna dalam pencariannya. Teori ini dapat dicapai dengan menggunakan DBpedia, satu set data RDF triple yang dapat menunjukkan hubungan antara konsep dalam suatu graf yang besar. Ia dibentuk berdasarkan maklumat berstruktur daripada Wikipedia. Mesin carian meta semantik yang dicadangkan ini menggunakan Yippy, satu mesin carian meta yang dapat menyusun hasil carian dalam bentuk kumpulan topik yang mungkin berkaitan sebagai rujukan. Mesin carian ini menggunakan Google untuk mendapatkan hasil carian dan menentukan maksud kunci perkataan pengguna dengan DBpedia. Akhirnya, ia memaparkan hasil carian dalam bentuk kumpulan topik seperti Yippy. Ia juga memberikan cadangan yang luas berdasarkan konsep yang berkaitan kepada pengguna. Eksperimen telah menunjukkan bahawa dengan menggunakan set data DBpedia, hasil carian daripada mesin carian yang bukan semantik dapat diperkayakan dengan unsur semantik sehingga tahap tertentu. Mesin carian yang dicadangkan ini juga dapat memberikan cadangan yang berkaitan berdasarkan kunci perkataan pengguna dalam pencarian.

## SYNOPSIS

*The difficulty of information retrieval on the web is proportional to the size of information available on the web. Nowadays information on the web has grown to a size at which even a good ranking algorithm could not produce a precise search result. The disambiguation that exists between several terms gives challenge on how a search engine should produce a search result since traditional search engines work based on pattern matching rather than the meaning of the term being queried. On the other hand, semantic search engines search for concept that focuses on the meaning of the input query, rather than considering a query as a group of string. However, the Web is still dominated by Web 2.0 in which information and data is presented in an unstructured manner and is only fit for human consumption. Hence, building a semantic search engine is a very challenging task and there are still a lot of improvements that needs to be done especially in field of Natural Language Processing (NLP) to achieve the desirable results. Hence, this research proposes a semantic meta search engine that utilizes the power of a traditional search engine and enriches the search result by trying to understand the meaning of the search query. This is achieved by making use of DBpedia, an RDF triple dataset that is derived from structured information on Wikipedia which gives a huge graph of how concepts are related to each other. The proposed meta search engine takes into account the design of Yippy, a meta search engine that has of the ability to cluster search result into a group of topics that may be relevant to the query. This semantic meta search engine also uses Google to generate the search results of a given query, explores the meaning of the query using DBpedia and finally displays the results in cluster similar to the technique used by Yippy. It also provides broad suggestions based on concepts that are related the query. The experimental evaluation showed that by using DBpedia dataset, the non-semantic search results can be clustered, rearranged and enriched with a certain degree of semantic. The proposed search engine is also able to provide suggestions on concepts that are similar to the query.*