A Framework of Universiti Kebangsaan Malaysia Patent: UKM Patent

Rosilah Hassan*, Abdullah A. Al-khatib* and Wan Mohd Hirwani Wan Hussain**
*Research Centre for Software Technology and Management (SOFTAM), Faculty of Information Science and Technology (FTSM),
**Graduate School of Business (GSB),
Universiti Kebangsaan Malaysia, 43600 UKM Bangi Selangor, MALAYSIA
rosilah@ukm.edu.my, khteb2003@gmail.com, wmhwh@ukm.edu.my

***Abstract*— A Meta-Search engine is an ideal tool that takes search expressions and sends them out to multiple search engines. It then returns the categorized results to the user. For example, we established a Universiti Kebangsaan Malaysia (UKM) Patent meta-search engine that takes patents from multiple search engines, such as Google patent and Patent Mall. This website allows UKM students to research patents inside Patent Mall, the UKM database and the Google patent search engine. This study focuses on web design and its development using several tools and programming languages, such as open source ASP.Net and SQL Server Database. A few of these are used to retrieve information from search engines, for example, the Google patent application programming interface (API) and Post Query to Patent Mall. In this project, we establish a new framework for the first meta-search engine that specialises in patents, which will minimize useless pages in search results. Additionally, the UKM database contains private university UKM patents. The UKM patent meta-search will extract (or retrieve) these patents depending on a Language-Integrated Query (LINQ) that is used for conveniently extracting and processing patents from the UKM database. Finally, the UKM meta-search is a new solution to the problem of obtaining requested information quickly.**

***Keywords*— UKM database, Google Patent, Patent Mall, API, LINQ**

I. Introduction

Searching for information is currently an activity of great importance. It searches large volumes of documents available to find those that best fit our needs in the shortest time possible. To this end, information search tools are implemented to help find information in a large corpus of documents. Therefore, several questions arise about these tools; especially about their performance and relevance of the results that they return. A meta-search engine is a search engine whose main feature is that it forwards a query to several other search engines simultaneously (as demonstrated in Figure 1) and collects and processes the results. The results can be easily arranged one after the other [1]. In other words, meta-search is software that draws its information from several other search engines. This allows users to enter the subject of their research once and access multiple responses from different search engines [2]. Some meta-search engines don't use an algorithm, but still present the resulted information of the sources found. The UKM patent meta-search is a friendly website that gives all similar patent information from other search engines. To achieve this goal, we need to be able to connect to it from this website. In this paper, the reason why the UKM patent adopts this meta-search engine is because it assists in generating results accurately and conveniently from an increasing number (millions) of patents and from many different search engines. Finally, the UKM patent meta-search engine allows researchers to quickly get accurate information from different search engines.

In this paper, first we reviewed the overview of meta-search engines and the benefits and limitations of them. The other parts of this paper are organized as follows: Section III proposes the framework for the UKM Patent meta-search engine, by presenting several possible methods and techniques to establish the UKM Patent meta-search engine website. Section IV shows the results and discussion of the framework and the website. Section V is the conclusion and recommendations for future work.

II. Meta-Search Engine

A meta-search engine is an internet search tool that collects search results using many different search engines, indexes, and clusters, and organizes them using cutting-edge technology. It then uses Latent Semantic Analysis to quickly discover the relevant results. It saves time using intelligent indexing to quickly focus only on the results of interest. Search engines have become a daily tool for internet users and the educational process. However, an increasing number of pages per day (in the millions) makes it difficult for researchers to find information. Therefore, they require an easier method to obtain information quickly. The UKM meta-search engine is a new solution for these problems. The architecture standard of meta-search engines is shown in Figure1.

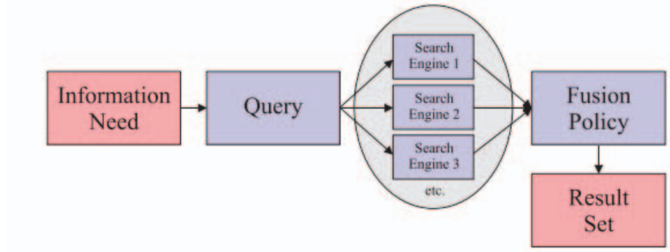

Figure1. The architecture standard of meta-search engine [6]

A. Limitations of search engines

According to [3], most existing webpages are not indexed. Each search engine captures a different percentage of the total; but no one can say exactly what proportion has been captured. The estimated size of the World Wide Web is at least 5,000 million pages; however, there is a much greater depth; estimated at about 500,000 million pages within databases whose contents are not captured by the search engines. These dynamic webpages take shape within a web server when a user asks for it; therefore, a conventional search engine cannot access them. Page United States Patent and Trademark Office (Patent Office and registered US marks) is an example: if a search engine can find your homepage, one can only search their database of individual patents by searching for the site itself. [4] The study result of a search engine query can sometimes give ambiguous results:

- Confusion between the part and the whole: Google can index some of the knowledge that is published on the web; but what is indexed by Google is not the entire Web, or all of the knowledge present worldwide.
- Confusion between quality and quantity: the countdown is preferred answers to the evaluation of the argumentative quality.
- Confusion between information and reality: the information provided by Google reflects reality without an intermediary.

B. Benefits of Meta-Search Engine

In [5], authors conducted an explicit study on the benefits of Meta-search engines:

- For access to multiple retrieval systems, the user must learn to work with only one interface. This may not take into account the differences in various search systems.
- Eliminate duplications in the search.
- From the user's perspective, it is more efficient if every query is entered only once. The query is evaluated in parallel and does not need to be entered for each system separately.
- Meta-search systems contribute to higher search completeness. Individual systems are not measured by the amount of indexed files only; but also by the content focus of the database. By utilizing meta-search engines so the user has a higher chance that their query will get more relevant documents.

C. Meta-Search Engine Architecture

Traditionally, the meta-search engine's server has to wait for responses from all search services to which it has forwarded the search, in order to start with the results representation [6]. This results in delays compared to a normal search engine. In order to counter this, a display that has been updated in each case, upon the arrival of different search results, can be carried out; or slow-response search engines can be excluded from the search. The current generation of meta-search engines also allows syntax translations; so that even more complex search queries can be sent to the respective search engines [7]. Two investigations on the problem of recovery of similar documents is identified [8], [9]. These works proposed fingerprint processes to represent the input document sets of relevant terms. Both architectures make use of meta-search engines to retrieve large lists of candidates for similar documents. In [8], authors use cosine similarity of the vector model to compare the search query with the documents snippets. In [9] text similarity algorithms such as Patricia and k-grams are used to compute the similarity. The following will study the standard architecture of a meta-search engine:

D. Standard Architecture

Apart from basic search engines, meta-search engines consist of four main software components; User Interface, Dispatcher, Display, and Personalization & Knowledge [10]. The software component architecture of a meta-search engine is shown in Figure 2.

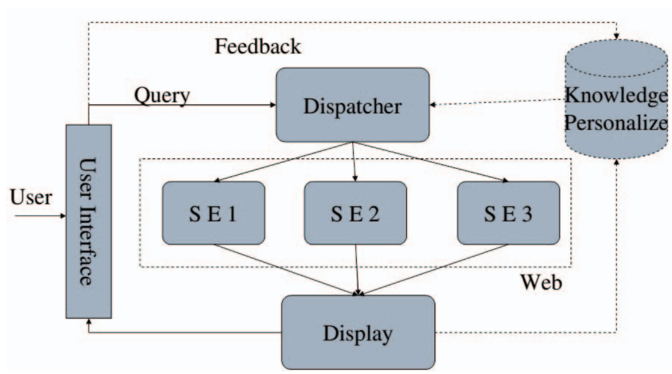

Figure2. Components of a meta-search engine [10]

1. ***User Interface:*** should be very easy to use. However, this makes no difference in providing a unification of the various interfaces of various search engines.
2. ***Dispatcher:*** deals with consultations that interact with different modules related to search engines. It is responsible for sending adequate consultations for each search engine and then collecting the returned results.
3. ***Display:*** generates a results page from the replies received. This may involve ranking, parsing and clustering of the search results or just plain stitching.

4. ***Personalization/Knowledge:*** may contain either or both. Personalization may involve weighting of search results/query/engine for each user.

III. Proposed Framework for UKM Patent Meta-Search ENGINE: System Architecture

The system framework (Figure 3) consists of components based on design layers developed in the UKM patent meta-search engine framework. Each component has proposed control of the tasks necessary to solve the problem of retrieving similar patents from the search engines. In this system, we have design oriented services based on Web Services. In this system, the service can combine results from other search engines. In this case, the services allow users to deal with the web interface to retrieve similar patents.

The interaction between the components is explained as follows. The process begins with text input at the user-interface and ends with a list of patents that are similarity ranked. The input is converted into a set of queries generated by the user-interface process that assigns greater probabilities of occurrence to the most relevant terms. Then, the queries are sent in parallel to a customizable list of search engines. Finally, the patents are retrieved, merged and ranked by the strategy proposed in the model and returned to the users; which, in this case, is the user interface. The following sections will describe the overview of the main components of the UKM patent meta-engine framework; which consists of four layers; front-end GUI, Middle-end meta-search, back-end search engines, and back-end UKM patent database. These components are illustrated in Figure 3.

A. Front-End GUI

The front-end GUI is the first layer in the system architecture of the UKM patent meta-search engine that interacts directly and transparently to the user with webpage interfaces. It is therefore merely a front-end that collects, organizes and manages information collection. However, it also translates user interaction in order to provide a text interface that is in charge of performing specific operations on the website, such as sending queries and displaying results on the front page of the website. This layer consists of two main components: results visualization Para-diagram and reformulation query. The results visualization is responsible classified by patents search engine, each in a tab. On each tab title patents documents are highlighted with a color for easy identification, the URL has a color that stands out, and left tab show the images of patent display in picture box. The user can click to image to get the URL of patents. The other component, reformulation the query provide a reformulation user query that generates a new query and initial request, in order to achieve more relevant documents from those provided by a non-reformulated query.

B. Middle-End Meta-Search

The Meta-search layer is the layer between the front-end GUI and the back-end. The interaction flow of this layer is either with the dispatcher or retrieved data that is saved from these search engines and the UKM database. The main task of the Meta-search in this system architecture is to distribute queries and merge the results retrieved from the search engines and the UKM database; all of the retrieved data that needs clustering and format conversion. There are five main components found in this layer; query dispatcher, result rank and merger, information extractor and sorting with filtering.

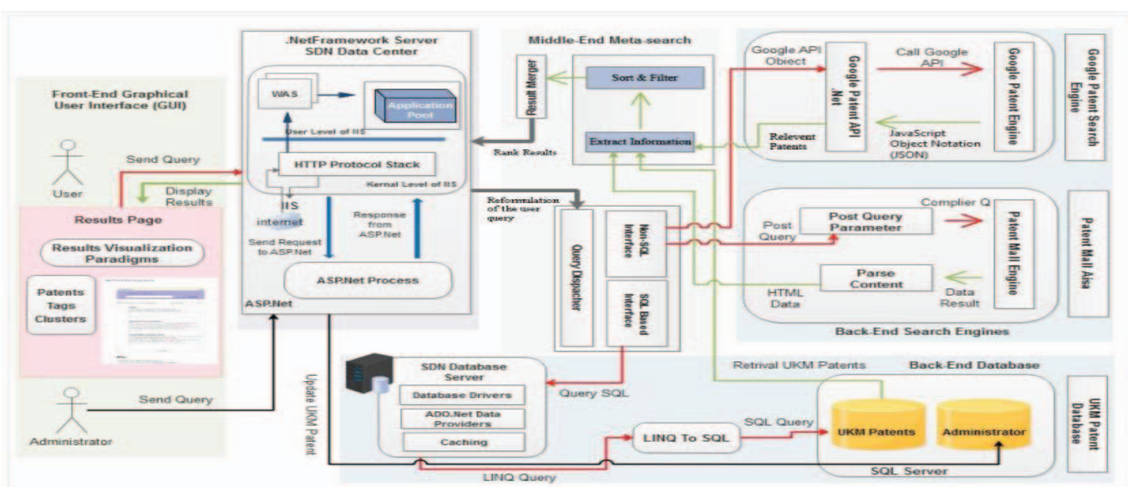

Figure 3. UKM Patent Meta-Search Engine Framework

The Query Dispatcher consists of two main components; SQL Based Interface and Non-SQL Interface. It is responsible route query to either Non-SQL or SQL, depend on it attribute. Each attribute processed by the Dispatcher to pass the query. The algorithm for rank and merge results use the Top Document search engine score (TopD) [10]. It allows classification and sorting of the retrieved patents. The extractor results from different search engines that need to be merged and stored either in a database or in xml format. This layer also sorts and filters the obtained results, based on the distance of cosines (commonly used in the vector model of information retrieval) [11], and neatly shows three tabbed results (Google patents, Mall Patents and UKM Patents).

C. Back-End

Back-end is comprised of two components as namely search engines and UKM patent database as shown in Figure 3. The search engine module in this system's architecture interacts with the middle-end layer through the query dispatcher, either directly to UKM database or indirectly retrieving information, like Google patents API and post query that supported by Patent Mall. However, there is no direct interaction from the back-end to the GUI layer.

The search engine back-end module is indirectly serves the front-end services; usually by being closer to the required resource or by having the capability to communicate with the required resource. For example, the resource in this system's architecture are the Google patent search engine and Patent mall databases. The main task of this layer is to collect patents from search engines using tools like API and Post query. API is a JavaScript library that allows inserting Google Patents into the UKM Patent meta-search website. It also obtains requests and query strings for patents that appear in URLs from Patent Mall. These patents consist of title, URL, images, and the description of the patent.

The UKM Patent database represents the established connection between the SDN Server and the databases. It is a standard that defines Microsoft SQL Server interfaces between middle-end and back-end servers on databases. The main task of this layer is to establish gateways that allow users to access different databases within a distributed environment using translation methods. It also the administrative center of the UKM patent meta-search engine website. The construction and implementation of the web, managing editing and insertion patent's content, are made through this elegant and intuitive interface. However, there is no direct interaction from the back-end to the GUI layer, but an indirect interaction through accessing the administration panel; either from the "Administration" menu item on the homepage.

IV. Result and Discussion

In order to assess the framework presented in Figure 3, a prototype system is developed and implemented using virtual C# Asp.net. The front-end user interface as shown in Figure 4, display the results obtained when running a query looking for "motor" in left side the results obtained from Google patent search engine is depicted. And the middle the results from Patent Mall display. In the right side the query results database from UKM. The figure shows that system is able to execute retrieving, processing, filtering and arrangement of the patents obtained have been performed. This interface displays classified patents from search engines, each in a tab. On each tab, the patent's title is highlighted in color for easy identification and a snippet (or text summary) is clearly displayed. We also test the query dispatcher to distribute the query to give quickly retrieved patents. We note that, the queries to external server takes 0.4 second more time to executed the connection through Google patent API take 0.2 longer time compare to access our local database patents. We have meat chairman of UKM patents in order to ensure usability of our proposed framework the user is satisfied with the system performance and usability since it gives quick access to UKM patent database. In addition, it was able to retrieve patent result from Google patent online service as well as mall patent. We hosted this project in the server support SDN to enable quick searches, above using a traditional hosting server.

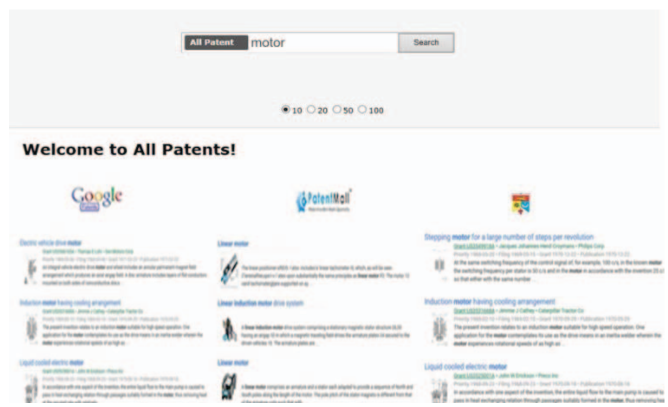

Figure4. Snapshot of UKM Patent Meta-Search Engine

Addition, we run 10 terms queries as shown in Table 1 the results prove that local query take less time to perform comparted to external source.

Table1. Results queries response

No	Query Terms	Response Time		
		Google P.	P. Mall	UKM P.
1	Motor	0.48	0.52	0.27
2	Computer	0.32	0.39	0.24
3	Network	0.44	0.62	0.21
4	Communication	0.56	0.52	0.18

5	Mobile	0.62	0.58	0.28
6	Laptop Computer	0.73	0.62	0.25
7	Barcode	0.45	0.48	0.19
8	System	0.49	0.68	0.21
9	Information	0.52	0.47	0.18
10	Information System	0.48	0.59	0.28

V. CONCLUSION AND FUTURE WORK

The UKM patent meta-search engine has become a necessity for students, professors and researchers, to give them patents easily and quickly. The goal is to enable users to find patents whose content meets their information needs. However, we found that the idea of relevance depends on the user's satisfaction on one hand, and different meanings carried by the terms of the application on the other. This finding is a weak point of looking for traditional information. It also represents the starting point for new research paradigms. After proposing this architecture, our future work will improve the rank and merge result's algorithm. In addition, we will implement our framework with new mechanism using optional filed on Internet Protocol Security (IPSec's) Encapsulating Security Payload (ESP) frame [12][13]. Based on [14] and use of Distributed Alternative Binding Cache mechanism (DABC), we will apply this system to accelerate retrieval results.

ACKNOWLEDGMENT

The authors would like to acknowledge the assistance provided by the Network and Communication Technology Research Group, FTSM, and UKM in providing facilities throughout the research. This project is partially supported under the ETP-2014-008.

REFERENCES

[1] Garman, N. (1999). Meta Search Engines. Online, 23(3), 74-78.
[2] Bassiouni, B. M. (2008). Meta search engines. Meta search engines, 4(2).
[3] Oberoi, I. S., & Chopra, M. (2010). Web Search Engines–A Comparative Study. Mälardalen University.
[4] Langville, A. N., & Meyer, C. D. (2006). A reordering for the PageRank problem. SIAM *Journal on Scientific Computing*, 27(6), 2112-2120.
[5] Sander-Beuermann, W., & Schomburg, M. (1998, July). Internet Information Retrieval-The Further Development of Meta-Searchengine Technology. In Proceedings of the Internet Summit, Internet Society, July (pp. 22-24).
[6] Glover J. E., Lawrence S., Birmingham P. W. and Giles C. L.; Architecture of a Metasearch Engine that Supports User Information Needs; NEC Research Institute, Artificial Intelligence Laboratory, University of Michigan; In ACM; 1999.
[7] MENG W.; Metasearch Engines; Department of Computer Science, State University of New York at Binghamton; Binghamton; 2008.
[8] Bilal Zaka. Empowering plagiarism detection with a web services enabled collaborative network. *Journal of Information Science and Engineering,* 25(5):1391{1403, 2009.
[9] Alvaro R. Pereira Jr. and Nivio Ziviani. Retrieving similar documents from the web. ´ J. Web Eng., 2(4):247{261, 2004.
[10] Jadidoleslamy, H. (2011). Introduction to metasearch engines and result merging strategies: a survey. *International Journal of Advances in Engineering & Technology*, 1(5), 30-40.
[11] Sidorov, G., Gelbukh, A., Gómez-Adorno, H., & Pinto, D. (2014). Soft similarity and soft cosine measure: Similarity of features in vector space model. *Computación y Sistemas*, *18*(3), 491-504.
[12] Al-khatib, A., Hassan, R., (2016, November). Case Study of Managing Computer Security and Analyzing Risk and Threat of an Organization at Seiyun University of Science and Technology, Yemen. *International Conference on Data, Internet & Education Technologies (DIET), 2017 2th International Conference*.
[13] Ahmed, A. S., Hassan, R., & Ali, Z. M. (2012, June). Eliminate spoofing threat in IPv6 tunnel. In Information Science and Digital Content Technology (ICIDT), 2012 8th International Conference on (Vol. 1, pp. 218-222). IEEE.
[14] Hasan, S. S., Hassan, R., & Abdalla, F. E. (2012). A new binding cache management policy for NEMO and MIPV6. *Journal of Theoretical and Applied Information Technology*, *36*(1), 113-117.

Dr. Rosilah Hassan works as an Engineer with Samsung Electronic Malaysia, Seremban, before joining Universiti Kebangsaan Malaysia(UKM) in 1997. She obtain her Master of Electrical Engineering (M.E.E) in Computer and Communication from UKM in 1999. In late 2003, she went to Glassgow, Scotland, for her PhD in Mobile Communication from University of Strathclyde. Her research interest is mobile communications, networking, 3G and QoS. She has been a senior lecturer at UKM for more than 10 years. She also specializing in Networking, QoS, IoT, Big Data, and Academic entrepreneurship.

Abdullah Abdulrahman graduated from Al-Ahgaff College Hathramout-Yemen in 2009. He works in communtiy college from 2009. He enrolled to study Master in University Kebangsaan Malaysia (UKM) in Computer Sceince (Network Technology) in 2015. His research interests are in Software Defined Network (SDN) and IP Security (IPSec).

Dr.Wan Mohd Hirwani Wan Hussain Ph.D is a university researcher at Graduate School of Business, UKM and internet marketing entrepreneur. He also specializing in technology transfer, commercialization technology and law. He has been involved in technology transfer activity and commercialization technology from university to the industry. He has published numerous articles in international conference and journal publication. Specialties: Technology Transfer, Commercialization University Research, Innovation Technology, Internet Marketing Strategy. Having over 8 years' experience in internet marketing, search engine marketing (SEM), search engine optimization (SEO), business development, product, service and commercialization technology.