Assisting Novice Researchers in Utilizing the Web as a Platform for Research: Semantic Approach

*Maizatul Akmar Ismail , #Mashkuri Yaacob , *Sameem Abdul Kareem & *Fariza Hanum Nasaruddin
*Faculty of Computer Science and Information Technology
University of Malaya
50603 Kuala Lumpur,
MALAYSIA
#Universiti Tenaga Nasional
Km 7, Jalan Kajang-Puchong,
43009 Kajang, Selangor,
MALAYSIA
*{maizatul,sameem, fariza}@um.edu.my, #mashkuri@uniten.edu.my

Abstract:
The access to internet has tremendously changed the way of information dissemination. The emergence of digital libraries and institutional repositories provides endless supplies of knowledge. Scholars in particular, make use of research output in the form of conference proceedings, journal and theses as references and guideline in generating new knowledge for the use of future generations. Novice researchers are the form of scholars which always drown in the ocean of information. This scholarly content resides on several heterogeneous databases that need to be integrated. Support in the early stage of study is crucial for novice researchers as it will give them some insights of where to seek for extra information based on institutions, people and research trend without having to go through tedious process of identifying this information all by themselves. Previous studies have identified support features that are useful for novice researchers among which are Relevant Literatures and Expert Detection. Thus, the purpose of this paper is to explore each of these support features and suggest state-of-the art development approach by utilizing Semantic Web technologies. The algorithms involved with each of the support features will be discussed. The result of the implementation shows significant information that can be utilized by novice researchers in accelerating research process. The evaluation of the information retrieved from the support features is elaborated. Future works in enhancing the proposed prototype are also discussed in the concluding remarks.

Key-Words: Novice Researchers, Support Features, Semantic Approach

1 Introduction

The web provides huge amount of information which offers unlimited resources for research work. On the other hand, it provide extreme amount of search results that may cause the scholars to be lost in Cyber Ocean. Support in the early stage of study is crucial for novice researchers as it will give them some insights of where to seek for extra information based on institutions, people and research trend without having to go through tedious process of identifying this information all by themselves.

This research aims to explore new approach to assist novice researcher in utilizing the web as a platform for scholarly activities. In this study, novice researchers are students who are just embarking in the research world. Few studies have defined novice researcher as someone who is 1) trying to be familiar with research processes before proceeding with the empirical study [1], 2) at the early phase of information gathering before they can put together the information as harmonious whole [2] and 3) trying to identify a research-worthy problem [3]. In our particular study, the chosen respondents are third year undergraduates and first year postgraduates from Faculty of Computer Science and Information Technology, University of Malaya.

The early stage in a research procedure as stated by Graziano and Raulin [4] i.e. idea generating phase, requires researchers to go through huge number of existing literatures. These literatures are stored in repositories such as databases and reside ubiquitously in the institution or organization that published the documents. Thus, a study to understand the current situation of users' utilization of academic resources, their role and information seeking methods for research purposes is needed. A questionnaire survey is administered to gather the information on whether students need extra features, more than what keyword-based search offers. A pilot study conducted on several postgraduate students reveals several issues regarding accessing information from open access portal, such as information are not up to date, dead links, available information are limited, and also some searching issue where the problem essence in constructing the right keywords. The study also shows that experienced users will use intuition and refine the search to get better results, whereas novice users will look at the retrieved result and "know it when they see it" which is time consuming and may lead to frustration. The survey questions are divided into six sections. The first section is concerned with the number of hours that students spent for research activity. This is to give the brief introduction to the student on the overall motive of this questionnaire. The second until fifth sections deal with the roles of students as readers/users and contributors of scholarly content. The final section is dedicated in identifying the students' methods of information seeking and also the motivating factors that will lead them to utilize academic repository for their research activity. The summary of the survey section is stated below:

- Utilization of academic resources for research purpose for undergraduates and postgraduates is low because students tend to use generic search engine such as Google and Yahoo to search for information which will result in the non-scholarly resources being treated as having an academic value.
  Solution: Provide easy-to-use system that allows the retrieval of scholarly documents with a single query.
- Students play a role as reader more than they are willing to play the role of the contributor.
  Solution: Provide a system that is beneficial for early-stage research work so that students are encouraged to be part of the contributor of the scholarly content in the future.
- Undergraduates request for resources that would fulfill their immediate needs. Majority of the undergraduates wish for software related resources such as sample of final year project, source code and open source software as it was part of the major requirement for their final year project. As there is no database available so far for those types of resources, there is need to create one for the benefit of future undergraduate students.
  Solution: Provide a database for undergraduate thesis project.
- The results from section three, four and six indicate that students need a common access to academic related resources to assist them in the research process.
  Solution: Provide an integrated scholarly database by utilizing a generic data model for scholarly communication.
- Majority of the undergraduates are motivated to use the Institutional Repository if it provides free access, user friendly interface and efficient search.
  Solution: Provide a system that would offer supporting features for early-staged researcher in identifying specific information in a particular research area. Students clearly indicate that a support environment that could assist their research activity is highly in need. Educational technology research concerned on how tools or computer can enhance the teaching and learning process (the research work in our context) by making a learning process more engaging, improving the needs of individual students, supplying access to a numerous information, and to encourage students to explore and create new knowledge [5, 6]. The full result and analysis from this survey can be found in Ismail et.al [7].

As this survey is about understanding the utilization of academic resources by the students and motivating factors that will encourage them to use the resources, it can be deduced that the level of academic utilization is still low due to several factors such as no such resources (such as repository for thesis and final year project software) are available and even though these resources exist, they are not integrated. The survey also reveals that students face difficulties in getting the resources in academic repository due to unavailability of support in the process of searching relevant information. The local students demonstrates similar behavior as stated in previous studies by Griffifth and Brophy [8, 9] except that postgraduates seem to appreciate the usage of scholarly resources for research work more than the undergraduates. The DEvISE project [10] has found that efficiency is most strongly correlated with users' general satisfaction which seems to "suggest that the amount of time and effort required from the user matters more than the relevance of the items found". On the other hand, both groups agree that three factors are useful in assisting users on the early stages of their research work. These are namely 1) Related Literatures, 2) Research Trend; and 3) Experts in Specific Research Area. The overall analysis of the surveys and the next research step is illustrates as follows (Figure 1):

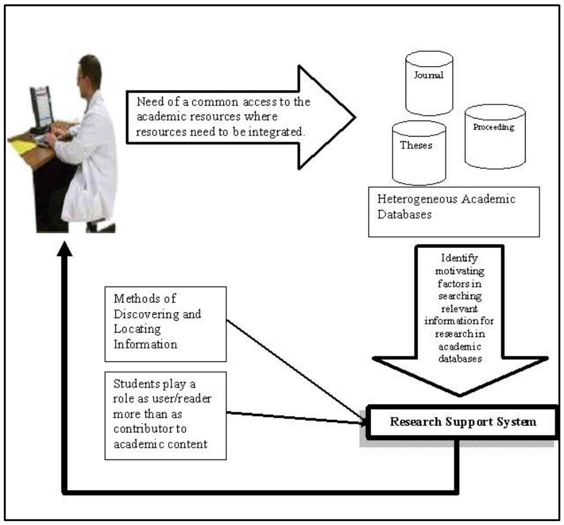

Figure 1. Diagram of Survey Analysis and Research Gap

Two important criteria derived from the survey are 1) the need to provide common access to information and 2) support features in searching relevant information for research works. Data in our study involved heterogeneous scholarly databases which are individually developed and separately maintained by different research group in FSCIT, UM. The production of academics activities from Faculty of Computer Science and Information Technology, University of Malaya (FCSIT, UM) are mostly reflected in the Malaysian Journal of Computer Science (MJCS), Malaysian Journal of Library and Information Science (MJLIS), locally developed Conference database and collection of theses stored in FCSIT DSpace Institutional Repository. Thus, a common access to this information is imperative since it provides valuable sources for research work. Common access refers to a standardized interface that allows user to access all the information that are "expressed using diverse vocabulary and inaccessible format" [11]. Semantic approach has been chosen as an integration method which have been described in [12, 13]. Figure 2 shows the semantic storage after the three databases (i.e. journal, thesis and conference) have been integrated.

After the data can be commonly accessed , the next step will be in exploring a new approach to assist novice researcher in utilizing the web as a platform for scholarly activities from the experienced researchers' perspectives. The results of the study can be found in Ismail et.al [14]

Support needed for research work have been discussed widely in [15], [14] and [16]. The envisage features in a system that facilitate research work is the one that enhance the quality of user interaction as stated in the Human Computer Interaction (HCI) principles [17]. This paper is going to discuss the methodology taken in realizing two of the support features identified in [14] by utilizing semantic approach.

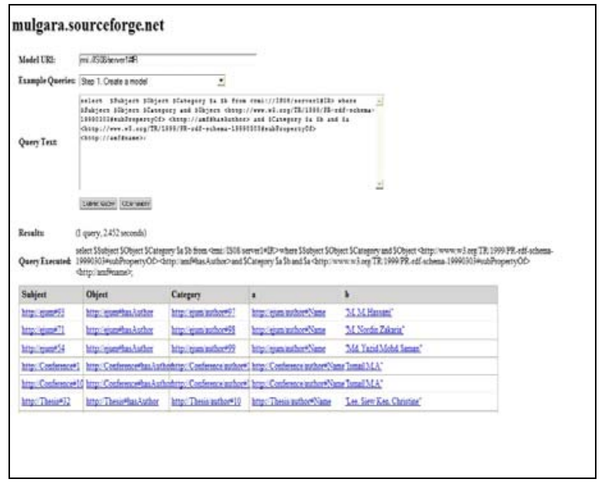

Figure 2. Integrated Scholarly Database

2 Methodology

2.1 Scholarly Documents Preparation

We utilized Ontogen, a semi-automatic ontology editor in assigning relevant topics to the scholarly documents such as thesis, conferences and journal articles [18]. Experts in four areas of Computer Science were asked to relate the documents with its suitable topics in the ontology. The topics ontology is constructed based on the research category in ACM and the categories that are suggested by the authors of the papers. For instance, the concepts and terms in the following table are extracted from the Malaysian Journal of Computer Science and Malaysian Journal of Library and Information Science (see table 1.0). These journals are chosen because of the high recognition received from international bodies for scientific community i.e. SCOPUS and International Science Index (ISI). Furthermore, these journals reflect the publication of high quality papers by majority of local researchers. The table shows the answer given by one of the expert based on the relatedness of topics in Information System (column) and Information Science (row). Relevant terms for each concept are supplied based on the keywords assign by the author of the documents.

Figure 3.0 in the next page visualizes the Ontogen's environment [18] where domain ontology for Information System and Information Science is constructed. Apart from that, Ontogen also supplies additional keywords based on the few algorithm i.e. k-Means and LSI which can be used to further enrich the vocabularies that describe the documents.

Table 1.0 : The cross relation of concepts from Information Science and Computer Science

Concepts	Terms	Synonyms (Equavalent Words)	Information System	Information Storage and Retrieval	Database Management
Computer Application in Information Science	Online Catalogs		X	X	X
	Portal Technologies		X		
	Common User Interface		X		
	E-commerce		X		X
	Internet Technologies		X	X	X
Digital libraries	Electronic Journals		X	X	X
	Electronic Resources		X	X	X
	Resource Assessment		X		
E-Publishing	Online information		X		
	E-Book		X	X	X
Information Management	Business Process Reengineering		X		
	Information Services		X		X
	E-Faculty		X		
	Electronic Records		X		
	E-Government		X		

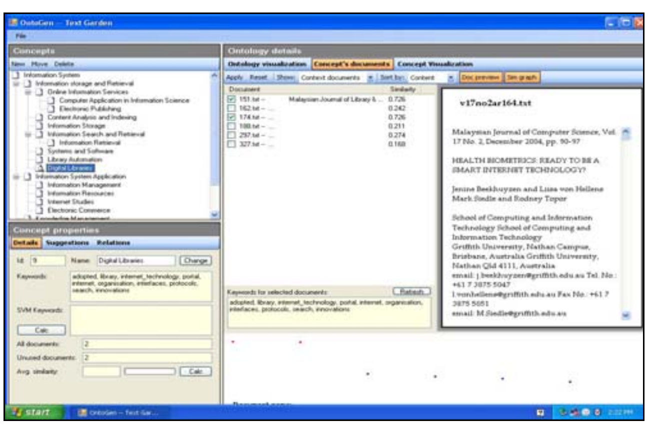

Figure 3.0 : Ontogen's environment in constructing domain ontology for Information System and Information Science.

2.2 Related Literature Module

Identification of relevant literature is identified by research expert to be the most important support that would assist novice researchers in the early days of their studies. In this study, a document is called technically relevant for a given search term if it contains the term in the title, abstract and author-assigned keywords. The relevancy of the documents is also determined by the document's attachment to the nodes in the given ontology. The Related Literature feature will retrieve information through a matching module which is supported by domain ontology such as ACM Computing Classification System (1998) and individually developed ontology based on research category suggested by the author. An ontology based semantic layer would enable the relationships between the various metadata schemes to be properly represented within the chosen ontology. The ability of using relationship information which is stored in ontologies enables semantic search engines to overcome the problems associated with existing search methods. The result of search will be the literatures based on those most related to the ACM category and Ontogen's constructed ontology with the entered keywords. This is beneficial because accurate categorization will provide quick content reference for the researcher and accelerate the progress of the search for related literature, as well as searches for the research on other online resources.

The algorithm that we applied for relevant literature module is as follows:

1. User Input Query
2. Extract Abstract, Topic and Keywords (User assigned keywords and Ontogen's suggested keywords) from triple store database.
3. Create an empty Array
4. Tokenized Abstract, Topic and Keyword (User assigned keywords and Ontogen's suggested keywords)
5. Add each Token to Array
6. Find number of user query in Array of Tokens
7. Rank and List Articles and Category based on frequency
8. Retrieve all other articles that are attached to sub-category and super-category of ontology
9. Display the result

2.3 Expert Detection Module

In detecting expert(s) in the field of CS and IT, the text corpus will be analyzed by using statistical technique to find specific experts in specific research field. Typically, these are some questions that novice researcher has in mind in order to justify other researchers' level of expertise:

- This author has any other papers?
- Who are the experts in digital libraries?

Resolving those queries manually would require further investigation and research which would require more time. In our work, in order to determine the expert, we use three heuristics as below (Table 2.0) :

Table 2.0 : Heuristics Used in Determining Experts

No.	**Parameters**	**Weightage (%)**
1	Number of Publications	50
2	Citation Analysis (Published papers in (1) that have been cited by others)	40
3	Sequence of Authors	10
	Total ((1) + (2) + (3))	100

Bibliographic measures are widely known in measuring the performance of the members of scientific community [19]. Numbers of publications in specific research area have long been recognized as indicators of expertise [15, 20, 21]. Thus, in determining the expertise, number of published papers is given more weight compared to other indicators that are described by Katz [21] i.e. citation of papers and citations per paper. Impact factors for published papers are measured by the citations that are received by particular papers [22]. There are few techniques in determining author impact factors. First, the h-indexed formulae is applied to measure the impact factors [23] : "A scientist has index h if h of his/her Np papers have at least h citations each, and the other (Np − h) papers have no more than h citations each." . For instance, h-index of 10 means that a researcher has published 10 papers where each had at least 10 citations. Second, cited references or citation analysis involves "counting how many times a paper or researcher is cited" in which concluded that " influential scientists and important works are cited more often than others" [24]. Google Scholar and Thomson ISI Web of Science are often used for citation analysis [25]. Thus, due to the coherently small size databases that we have as data test beds, we chose citation analysis in deciding the factor that determine one's expertise. Consequently, we assigned citation analysis weight lower than the number of published papers i.e. 40%. Other indicator of expertise that is known by the academic community but seldom given any attention or consideration is the order of authors' appearance in academic papers. Subramaniam [26] stated that "The question of ordering the names of coauthors is highly complex and elusive, while it is generally true that the name of the principal investigator is almost always mentioned first… "(p.36). Even though the name of co-authors are sometimes arranged according to alphabetical order, the principals investigators name whom contributed the most for the research publication is always placed at the beginning [27]. Thus, we take this heuristic into consideration and give the weightage of 10% in determining expertise in particular research area. As Subramaniam, [26] observed, it is not uncommon to find ten or more names for research publications, random analysis on ACM and IEEE publication from the year 2000 onwards shows that the most authors in a single publication are nine. Therefore, we put the maximum weight for the first author weight and the least weight for the last author. However, to further verify the weightage given to each of the factors, we cross examine the factors with the research expert and ask them to rank and weigh each of the factors. The result of the final analysis is integrated inside the coding.

Table 3.0 below shows the research experts' view the factors that determining expert based on research publications.

Table 3.0 : Factors that determined expertise based on research publications.

No\factors	**NP**	**CA**	**SA**	**RT**	**CW**	**FJ**	**IJ**
E1	X	X	X				x
E2	X	X	X				
E3	X	X	X	x			
E4	X	X	X		x	x	x
E5	X	X					
E6	X	X	x				x
E7	X	X	x				
E8	x	x	x				

Legend:
E1 – E8 : Research Experts
NP: Number of publications
CA: Citation Analysis
SA: Sequence of Authors
RT: Relevancy of topic of publications with the scope of journals
CW: Collaborative work with authors from other institutions
FJ: Funded Research project
IJ: Impact factor of Journal

We have asked the research experts to put a weigtage for the factors suggested with the highest weight for the most important factor and lowest weight to the least important one. The total count of weight for the factors must be 100%. From the survey, we found that number of publications, citation analysis, sequence of authors and journal impact factors are the factors that are chosen by majority of the experts. Thus, we will incorporate these factors for analysis in determining experts based on research publication. However, the impact factor for journal is only applicable to journal type of publications which are the MJCS and MJLIS. Impact factor is the "average number of times published papers are cited up to two years after publication" [28]. Conversely, considering the fact that MJCS and MJLIS are new journal and the impact factor are yet to be determined, we put more weight to the authors that have journal as part of his/her publication since our databases consist other type of publication such as conference articles and theses. Other factors are disregard because they are chosen the least with average weight of less than zero.

Based on the weightage given by the research experts, we calculate the average weight for each factor as shown in Table 4.0:

Table 4.0 : Redefined Heuristics Used in Determining Experts

No.	Parameters	Weightage (%) in Average
1	Number of Publications	38.1%
2	Citation Analysis (Published papers in (1) that have been cited by others)	37.5%
3	Journal impact factor	19.0%
4	Sequence of Authors	5.4%
	Total = ((1) + (2) + (3) + (4))	100%

Journals have always been regarded as the most prominent indicators of research experts. On the other hand, research progresses rapidly in the area of Information Technology and Computer Science. As new technology and improvements to current knowledge are discovered constantly; conference articles and thesis are still considered as relevant in determining one expertise.

Researchers often involved in various research works. For this, we also suggest the super-category based on ACM's and Ontogen's classification with the assumption that the super-category of research topics will give the user some idea on the people who are working in similar area of research. We take into account the nearest subcategory and supercategory of research area in ranking other experts who work in similar area of research. Figure 4.0 below shows the ontology Information System's categories.

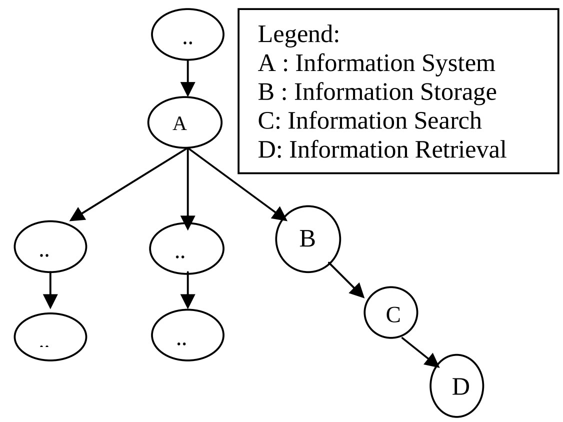

Figure 4.0: Information System's Ontology

For instance, if the user entered Information Retrieval as query and want to see the experts working in this area, we applied the four heuristics that we have mentioned above (see table 4.0). We also suggested other experts that work in similar area based on the supercategory of Information retrieval and rank these experts according to the distance of nodes. The depth of nodes in our data is four, thus we assign most weight to the nearest nodes and least weight to the nodes that are far from the user query. As the research topics are suggested by the research experts in the category, this technique is considered reliable because novice researchers can have more options in selecting other expert researchers in guiding them in the research work in case that the chosen topics do not have any experts in it. The algorithm that we applied for expert detection module is as follows:

1. User Input.
2. If input query exist inside ontogen files
    A. Go to Step 3
    B. Else Go to Step 18
3. Search inside ontology for the articles in the Query category.
4. Extract the articles from triple store database.
5. Keep result in the array
6. Find the ID of the articles to search its title.
7. If Array != NULL
    A. Go to Step 7
    B. Else go to Step 8
8. Extract the titles.

9. Search for that title in the Reference database to see who cited it & if it is cited more than two times – assign 0.375 weights (1).
10. Check Sequence of Authors (2) -
assign weight (range 0.09 to 0.1)* 0.054
12. Extract the authors of those Articles.
12. Keep Authors in an array and assign 0.19 if the articles are journals (4)
13. If Array != NULL
    A. Go to Step 13.
    B. Else go to Step 11.
14. Compare the authors to find which
one repeated more.
15. Store the authors name and articles ID and (3) number of repeated terms in the array – assign 0.38.
16. Sort using (1) + (2) + (3) + (4)
17. Show List of articles.
18. Find tree of the nearest super category of user input inside ACM CSS.
19. Display the result.

For the prototype implementation, Java Server Page (JSP) programming language, Apache Server and Mulgara Semantic Database are utilized. Figure 5.0 depicts the screenshot of the front page of prototype system, Research Support System (RSS).

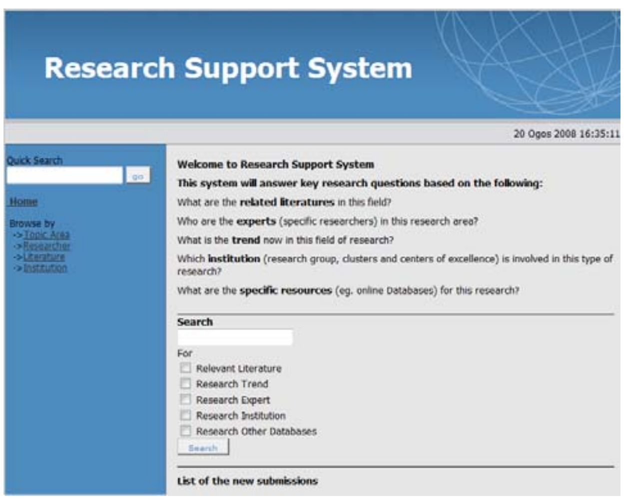

Figure 5.0: Screenshot of the front page of the system prototype

User can choose from the five modules that are available. However, in this paper we will just concentrate on two modules which are relevant literature and research expert. To use the system, user need to enter keyword(s) and click one of the modules. For instance, if the user want to find the relevant literature for the keywords “database management”, he/she only need to type in the keywords inside the textbox and click the relevant literature module checkbox as shown in Figure 6.0 below :

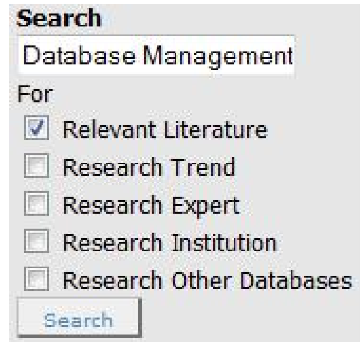

Figure 6.0: Screen for User Input Query

After user click on the search button, a web page that displays the results of the query which ranked by frequency of terms will appear as shown in Figure 7.0 below:

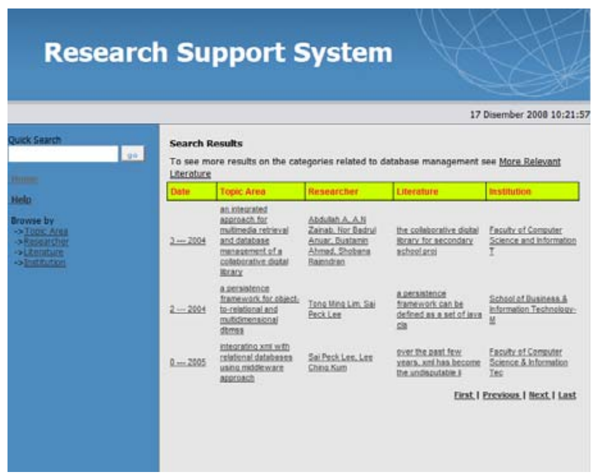

Figure 7.0: Results for Related Literatures for Database Management

User can also find out the relevant category to database management by clicking at “More Relevant Literatures” text. The result displayed for this page (in Figure 8.0) are based on the ontology on which the articles are preprocessed in the scholarly document preparation stage as discussed in subsection 2.1.

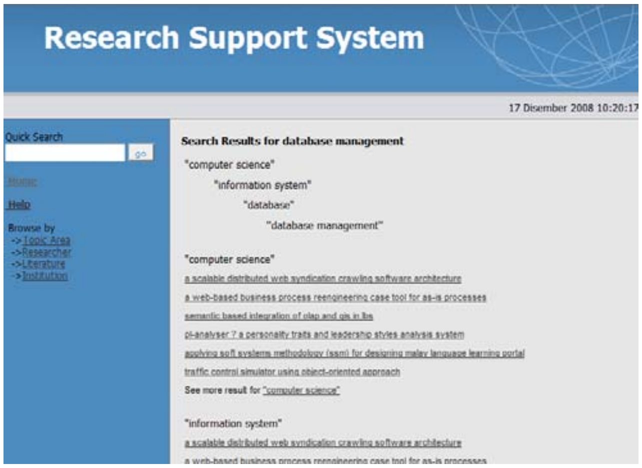

Figure 8.0: Results based on Ontology for the term Database Management.

For Expert Detection module, user will be presented with the name of experts in particular research area as indicated in the user's query. For instance, Figure 9.0 below depicts the suggested experts for the "Knowledge Management" research area:

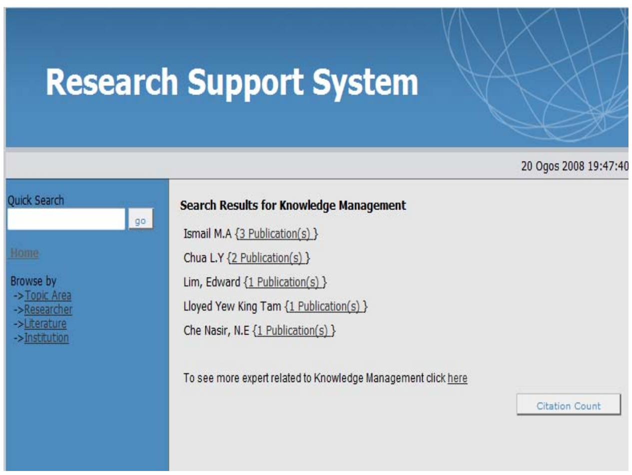

Figure 9.0: Results for Experts in Knowledge Management research area.

3 Evaluation

Precision and Recall strategy will be used to measure the performance of Information Retrieval (IR) of the system prototype. Recall is the proportion of retrieved elements among the existing relevant ones (which should have been retrieved). Precision is the proportion of relevant elements among the retrieved elements by that search [29]. The aim of this system is to enable higher Precision score with low loss in recall. A good IR system should retrieve as many relevant documents as possible which means having high recall, and it should retrieve very few non-relevant documents which means having high rrecision. Unfortunately, these two goals have been quite contradictory over the years and the techniques that tend to improve recall tend to hurt precision and vice-versa [29].

$$\textbf{Recall=} \frac{\text{Total Number of Relevant and Retrieved documents}}{\text{Total number of results Relevant}}$$

$$\textbf{Precision=} \frac{\text{Total Number of Relevant and Retrieved documents}}{\text{Total number of results Retrieved}}$$

Electronic Journals of University of Malaya (EJUM) provide two type of search: Basic and Advance. Basic search enable user to fill in keywords for a particular field such as Title, Authors, Keywords, Abstract, Journal Classification and Category. For Advance Search, the results are all the journals which have similar or matching string inside the documents. Studies by Zainab et.al [30] on the usability of EJUM stated that keywords search under the Basic search options are opted by majority of the EJUM's user whereas very few opted for advance search. Taking this point as consideration, the design of RSS corresponds to the basic search strategy in EJUM yet equipped with advance semantic features enabled by the usage of ontology. Furthermore, the user of RSS, whom are the novice researcher in the field of IT and CS are competent in searching due to the exposure to the Internet at the early staged of their study in bachelor degree. Studies by Tenopir and King [31] shows that the increase level of searching competency leads to the behavior of searching online journals more as opposed to just browsing. Study by Huzaimah [32] indicates that Electronic Journal of University of Malaya (EJUM) users face problems while searching and browsing for related articles. This might be due to the fact that users could not find the information needed or do not search for the right terms. Huzaimah's study suggested that the search function should be simplified. This could be related to the basic and advance search features of EJUM. As a little technical knowledge is needed to place the keywords in the right place, users are perceived to have difficulties in getting the information that they want due to the lack of supporting features of the search function. Other problems may be due to the inaccurate use of keywords. The appearance seems to be more "literature focused" and do not provide extra information other than what is related to the article itself. In the other words, there is lack of analysis on the results presented.

To evaluate the result of document retrieval by both EJUM and RSS, we will use the recall and precision strategy. For Basic Search in EJUM, five articles are retrieved as results from the user query of "database management" which are indicated in table 5.0 below:

Table 5.0: Comparison of Precision and Recall between RSS and EJUM

Query Terms	**RSS**	**EJUM**
Database Management	P=9/11=81% R=9/10=90%	P=5/5=100% R=5/10=50%
Knowledge Management	P=4/10=40% R=4/9=44%	P=6/6=100% R=6/9=66%
Pattern Recognition	P=13/14=92% R=13/19=68%	P=12/12=100% R=12/19=63%
Software Construction	P=12/14=85% R=12/12=100%	P=0/0=0 R=0/12=0
Distributed System	P=4/4=100% R=4/4=100%	P=4/4=100% R=4/4=100%
Average	P= 79.6% R= 80.4%	P=80% R=55.8%

The table above shows the result of recall and precision for the query of database management. As RSS provides ontology-based (and graph) search to find all the relevant data (from the string matching keywords and expert's assigned ontological nodes), the recall measure would be maximum. In this case for instance, the term "database management" yielded three results based on keywords match, six results from the documents that are attached to the concept of database management and three results from the documents attached to super category of Database Management i.e. Database. The precision value for RSS is 100% since the total number of relevant document match the total number of retrieved documents. This shows that while the system provide high recall (as compared to EJUM), it also maintains high level of precision. The performance of the algorithm is said to be better if both measures are close to unity [33]. The average value of recall and precision for the relevant literature module in RSS exhibit the better performance of semantic type of search compared to EJUM. On the other hand, results from EJUM demonstrate the inverse relationship between the recall and precision value which is normal in information retrieval concept as the increase in precision value hurt the recall.

5 Conclusion

Most of the important steps in identifying the right information at the right place could be done faster if research students have been told to do the "right" things first. Certainly, it does depend on the students' effort to ascertain the success of their research. Thus, extra support on the earlier stage of the study would definitely be helpful. This paper highlight two types of support features needed to assist novice researcher and the steps taken in realizing it. Future work involves the development of other features identified in [14] i.e. Trend Detection module, Relevant Online Database module and Relevant Research Institution module. We will conduct a study to evaluate the usability of the prototype system. As an important dimension in the evaluation of a virtual learning/research environment is the assessment of e-learner satisfaction [34], a questionnaire survey based on Software Usability and Measurement Inventory (SUMI) would be the best solution for study instrument.

References:

[1] Backman, K and Kyngäs, K.A (1999), Challenges of the grounded theory approach to a novice researcher, Journal of Nursing and Health Sciences, Vol 1 No. 3, pp 147-153, ISSN 1442-2018

[2] Ausband.L (2006), Qualitative Research and Quilting: Advice for Novice Researchers, *The Qualitative Report* Vol 11 No 4 ISSN 764-770I

[3] Ellis, T.J and Levy, Y. (2008), *Framework* of Problem-Based Research: A Guide for Novice Researchers on the Development of a Research-Worthy Problem, Informing Science: The International Journal of an Emerging Transdiscipline (Ed) Volume 11

[4] Graziano, A.M, and Raulin M.L, (2000) , Research Methods: A Process of Inquiry, 4th Edition, Allyn and Bacon, Boston.

[5] Perivoliotis, M.C, Ioannou, H. and Toumazatou, B. (2005a), Modern Technology and New Educational Systems : A Hellenic Design Case Study, In the Proceeding of WSEAS International Conference on Engineering Education, 8th – 10th July 2005 , Athens, Greece

[6] Perivoliotis, M.C (2005b), Adapting Distance Learning in a European Module, WSEAS TRANSACTIONS on Advances in Engineering Education, Vol 2 No. 2, pp 92-100, ISSN 1790-1979

[7] Ismail M.A, Mashkuri Yaacob and Sameem Abdul Kareem (2008d), Research Activity on the Web: Role of Students and Searching Needs in a

Malaysian Local University, International Journal of Learning, Vol 14, ISSN 1447-9494, pp 11

[8] Griffiths, J.R. and Brophy. P. (2002). Student searching behaviour in the JISC Information Environment." Ariadne Issue 33

[9] Griffiths, J.R. and Brophy. P, (2005). Student Searching Behavior and the Web: Use of Academic Resources and Google." Library Trends 53(4):539–554.

[10]Johnson, F.C., Griffiths, J.R and Hartley, R.J (2001). "Devise : A framework for the evaluation of Internet Search Engines." In Library and information commission research report CERLIM (Centre for Research in Library and Information Management) 0147

[11] Falbo, R.A, Guizzardi, G. and Duarte, K.C, (2002) "An Ontological Approach to Domain Engineering." In International Conference on Software Engineering and Knowledge Engineering. Ischia, Italy ACM

[12] Ismail, M. A., Yaakob, M. & Kareem, S. A. (2008a) Integrating Scholarly Content: Semantic Approach. IN IEEE (Ed.) *5th International Conference on Information Technology and Applications (ICITA 2008).*Cairns, Queensland, Australia.

[13] Ismail, M. A., Yaakob, M. & Kareem, S. A. (2008b) Integration of Heterogeneous Relational Databases: RDF Mapping Approach. IN UKM (Ed. *International Symposium on Information Technology (ITSym08).* Kuala Lumpur, Malaysia, IEEE.

[14] Ismail, M. A., Yaakob, M. & Kareem, S. A. (2008c) Semantic Support Environment for Research Activity. *Journal of US-CHINA Education Review,* 5**,** 36-51.

[15] Kampa, S. R. (2002) Who are the Experts? E-Scholars in the Semantic Web. *Department of Electronics and Computer Science.* UK, University of Southampton.

[16] Ismail, M. A., Yaakob, M. & Kareem, S. A. (2007) Semantic Search Engine in Institutional Repository : An Ontological Approach. *Proceedings of the International Conference on Libraries, Information and Society ICoLIS 2007.* Kuala Lumpur -Malaysia.

[17] Moroz-Lapin, K (2008), Active learning in the education of human computer interaction, WSEAS TRANSACTIONS on Advances in Engineering Education , Vol 5 No.8, pp570-579, ISSN: 1790-1979

[18]Fortuna, B., Grobelnik, M. & Mladenić, D. (2006) Semi-automatic Data-driven Ontology Construction System. *Proceedings of the 9th International Multi-Conference Information Society.* Ljubljana, Slovenia.

[19] Katz, J. S. & Hicks, D. M. (1997) Bibliometric Indicators for National System of Innovation. *IDEA Paper Series.* STEP Group.

[20] Kampa, S. & Carr, L. (2000) Web Scholars. *9th International World Wide Web Conference Amsterdam.*

[21] Katz, J. S. (2000) Scale independent indicators and research evaluation. *electronic working paper series.* Brighton:SPRU.

[22] Yinian, G. & Zainab, A. N. (2001) Publication productivity of Malaysian Researchers in the Field of Computer Science and Information Technology. *Malaysian Journal of Library and Information Science,* 6**,** 1-23.

[23] Hirsch, J. (2005) "An index to quantify an individual's scientific research output,. *"Proceedings of the National Academy of Sciences.*

[24]HEALTHLINKS (2008) Impact Factors. Washington, University of Washington Health Sciences Libraries.

[25] Harzing, A.-W. (2008) Google Scholar **-** a new data source for citation analysis. . University of Melbourne.

[26] Subramaniam, K. (1983) Bibliometric studies of research collaboration: A Review. *Journal of Information science,* 6**,** 33 - 38.

[27] Zuckerman, H. A. (1968) Patterns of name ordering among authors of scientific papers: A study of social symbolism and its ambiguity. *American Journal of Sociology,* 7**,** 276-29

[28]Journal Impact Factor (2008) at url http://www.sciencegateway.org/impact/

[29]Kobayashi, M. & Takeda, K., (2000), "Information Retrieval on the Web", *ACM Computing Surveys*, vol. 32, no. 2, pp. 144-173

[30] A.N. Zainab; Huzaimah Abdul Rani & T.F. Ang. Using journal use study feedback to improve acessibility , Electronic Library, Vol. 25, no. 5, 2007: 558-574. ISSN 0264-0473

[31] Tenopir, C and King, D.W (1998) Designing the Future of electronic journals with lesson learned from the past: economic and use patterns of scientific journals, In the Proceeding of the IEEE Advances in Digital Libraries Conference, Santa Barbara, CA, p.11-17

[32] Huzaimah Abdul Rani. (2006). Academic usage of electronic journals: focusing on EJUM (Electronic Journal of University of Malaya. MLIS Thesis, Faculty of Computer Science and Information Technology, University of Malaya, Malaysia

[33] Hoang, L.M (2006), Emerging Trend Detection from Scientific Online Documents, PhD Thesis, Japan Advance Institute of Science and Technology, p41

[34] Katsidis, C, Anastasiades P.S, and Zacharopoulos P.G, (2008), e-learning at the Technological Educational Institute of Crete: An Evaluation Based

on the Student Experience, WSEAS TRANSACTIONS on Advances in Engineering Education, Vol 5 No.8, pp529-538, ISSN: 1790-1979