Semantic support environment for research activity*

Maizatul Akmar Ismail[1], Mashkuri Yaacob[2], Sameem Abdul Kareem[3]

(1. Department of Information Science, University of Malaya, Kuala Lumpur 50603, Malaysia;

2. Department of Computer System and Technology, University of Malaya, Kuala Lumpur 50603, Malaysia;

3. Department of Artificial Intelligence, University of Malaya, Kuala Lumpur 50603, Malaysia)

Abstract: Scholarly activities are a collection of academic related activities such as research, teaching and consultation work which result in research outputs such as journals, theses and articles in proceedings. The output will then be disseminated to researchers all over the world by means of the WWW. The four pillars of this scholarship i.e. discovery, teaching, application, and integration of knowledge are supported by research activities which often start with the process of identifying key research questions. The ocean of information available on the internet could sometimes cause the scholars to be "lost in Cyber Ocean" Thus, this study aims at recognizing the key research questions and suggesting a support environment for scholarly activities for novice researchers. A structured interview was conducted among researchers to capture three pertinent aspects of research activities: becoming proficient in the new field of research, reviewing a research paper and identifying special research interests in specific research areas. The analysis of the interview shows pertinent questions that scholar make while carrying out their research. The feedback provides an overview of the range and altitude of support needed by scholars in seeking information especially in the electronic journal environment. Based on the result of the study, a proposal for system prototype and future work will be discussed.

Key words: research questions; researchers; Malaysia; semantic support environment

1. Introduction

Over the last two decades, the access to internet has tremendously changed the way of disseminating information to the end user. Scholars in particular, are among the users who benefit from the advent of World Wide Web (WWW) where the process of seeking and disseminating information has become easier. The huge content of web resources provides an endless supply of knowledge. The emergence of digital libraries and Institutional Repositories (IR) contribute in populating the web with research material. Institutional Repository became a free platform of knowledge sharing for researchers around the globe.

Boyer's (1990) view in determining academic excellence focus on four aspects of scholarship i.e. discovery, teaching, application, and integration of knowledge. Weiser (1996) pinpoints three basic forms of scholarship (as

* Acknowledgements: This research was supported in part by the Fundamental Research Grant FR198/2007A and FP096/2007C. Special thanks to Sheyda Shahidi and Hero Modares for their assistance. The earlier version of this paper is presented in the 1st Malaysian Educational Technology Convention, 2007. Many thanks to all the people that contribute directly and indirectly towards this research.

Maizatul Akmar Ismail, MSc, Department of Information Science, University of Malaya; research fields: information retrieval, semantic web.

Mashkuri Yaacob, Ph.D., professor, Department of Computer System and Technology, University of Malaya; research fields: discrete wavelet, operating system, integrated information.

Sameem Abdul Kareem, Ph.D., associate professor, Department of Artificial Intelligence, University of Malaya; research fields: neural network, medical intelligent system.

stated in Virginia Polytechnic Institute and State University's Website, 2007), namely:

(1) *Discovery* of new knowledge;

(2) *Development* of new technologies, methods, materials, or uses;

(3) *Integration* of knowledge leading to new understanding.

The three pillars of this scholarship are supported by research activities that often start with the process of identifying key research questions. Further information is solicited from discussions with peers and colleagues, attending workshops and relevant conferences, reading lots of literatures and making intelligent judgments in order to sort out what is relevant and what is not. Scholars involved in advance learning need to identify key research areas in order for them to be proficient in their research fields. These scholars usually have to undergo the tedious process of locating the right information and hence need the most support in the process of information seeking during the early period of their research. Most academic literatures repository such as Institutional Repository are sort of "disconnected" from other important information such as expert in particular research area and trend of research topic in the scholar market. Users such as final year undergraduates and early staged postgraduates are the group that would benefit from a little help from this repository if it could provide extra information apart from the literature itself. This early stage is crucial for novice researchers as it can give them some insights of where to seek for extra information based on the institution, people and research trend without having to go through the tedious process of identifying this information all by themselves. Consequently, a scholarly support environment is needed to be embedded within an academic repository to assist users locating the "relevant" and the "right" information especially for novice researcher who is in the stage of finding the right institution and research supervisor.

Finding relevant literature is also a crucial step in research activity. One study of Huzaimah (2006) 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 reason 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.

Graziano and Raulin (2000) averred that research or simply search again is "*a systematic searching for information*", involving a series of inquiring procedures and activities. "Systematic" here refers to the structured sequence of procedures which govern the research process. "Searching" could be defined as an act of trying to find information that is related to one's need. Guha, et al (2003) further categorized the searching activity into two:

(1) Navigational search. In this type of search, user will supply the search engine with word or phrase which he wishes to find in the document. For instance, user might supply the combination of words such as "session 2 presentation ICOLIS" in order to obtain a particular document or web page. The query does not provide any concepts or meaning and the user's intention is to merely find the document that contains those words.

(2) Research search. The user will supply the search engine with a word or phrase about which he intends to find more information. The user might have a clue or no clue at all on the type of documents that he will retrieve as a result. More exactly, the user is making an effort to collect a few documents that will finally give him the

information that he want. For instance, the user might supply a phrase such as “knowledge reuse” to search the collection. The query provides meaningful concepts that are related to other concepts such as knowledge-based classification, knowledge sharing, ontology extraction, problem-solving methods etc. The queries will not only yield results based on keywords but also documents which contain those concepts. This is the search type that we aim to improve with using the technologies of Semantic Web.

Semantic Web is an intelligent web that uses some languages and standards to retrieve and process the information on the web. Semantic Web technology have further extended the capabilities of current WWW by providing access to the “deep web” in which the digital content are not directly accessible by generic search engines (Guha, et al., 2003). Alas, the study on utilizing the semantic approach on specific database (such as EJUM) is still lacking and there maybe room for improvement. Most of the academic repositories are too “literature oriented” which only provides information related to the metadata that describes the article such as title, authors, keywords, abstract, category and year of publication. On the other hand, novice researchers need to identify other related research factors such as experts in particular area and research trends that might help them in accelerating the research process. In the case of domain specific collections such as EJUM, the querying of information by the user requires knowledge of the technical terms of the domain and the structure of specific databases. At the moment, the collection can be queried using key word(s) which utilizes the string matching technique.

This particular research intends to explore a new approach to assist novice researcher in utilizing the web as a platform for scholarly activities from the researchers’ perspectives. This research also aims at improving the access to the “deep web” data by semantically augmenting EJUM searching capabilities through the utilization of an ontology and Resource Description Framework (RDF) descriptions in order to enhance the system’s search capabilities and providing other support features especially for novice researchers. Novice researcher, in our context is defined as someone new to a research activity.

This paper will discusses the related work in conducting research activities. Next, the research methodology which involves the interview with experienced researcher in gauging their perception in carrying out usual research activities is presented. The section following this discusses is the results of the interview. The analysis section will further analyze the results in determining the supporting factor needed for research work, which is then followed by description on the prototype system which consists of supporting features in academic repositories. Finally, we describe the conclusion and future work.

2. Literature study

Research activities aimed at identifying new area that can be further explored, thus contributing novel ideas to form a new knowledge or add to the existing body of knowledge. It requires the researchers to have the ability to do detective work (Kampa, 2002) such as identifying seminal papers in “ontology construction” as an example and other prominent researches in the area. Such information is quite difficult to be obtained as it would involve a series of investigative work in order to determine the specific information. Thus, three pertinent aspect of research activities are: becoming proficient in the new field of research (by asking research questions), reviewing a research paper and identifying special research interest in a specific research area which is then chosen to be further studied.

Graziano and Raulin (2000) stated that there are seven stages in conducting research work. The first phase that is the idea generation is the most important phase as it will determine the direction of the research. Most of

the time researchers would proceed with research topic that would interest them and cause a burning desire for them to seek for the answer. Thus, it is crucial at this stage for the researchers to see what is there (e.g. relevant literatures, institution that excel in that particular research work and also the research trend) for them to excavate further. Next, problem definition phases which concerns on the refinement of the ideas develop in the first phase. The ideas need to be supported with the mounting demand for the problem to be solved with new techniques and advanced technologies. This will involve thorough investigation on past literatures on the existing techniques and their limitations. The loopholes of previous study will open a door to a new research topic. The third phase i.e. the procedures-design or the research methodology phase. It involves a series of carefully plan steps in order to achieve the research objective(s) and solution to the problem defined in the second phase. The fourth phase is observation or implementation that deals with carrying out procedures specified in the third phase. Research work in Computer Science and Information Technology would normally involve with experimental and development types of research. The data gathered in this phase will later be analyzed and interpreted in the next phase. Next, the fifth phase is the data analysis phase that deals with scrutinizing the data gathered in the previous phase and chooses suitable data analysis technique. Statistical analysis is the most popular technique used to analyze data, which can provide variety of comparison method to suit the types of data gathered by researcher. The sixth phase is the interpretation phase that involves with making intellectual justification on reason why such result is obtained. This phase is also important as the new findings will contribute to the existing body of knowledge and help answering the research question stated in the first phase. The final phase is communication that concerns with disseminating the findings of study to the scholarly community world wide by presenting it to seminar and conferences. The publication out of research work provides continuous research cycle as weaknesses and loopholes reported in one research would open an opportunity for a new research work. The seven phases of research procedures introduced by Graziano and Raulin (2000) are very much related to the research questions posed in the research activity. Each phase is interconnected with the three suggested scholarly activity where it involved investigative and detective works in examining the relevant literature for each of the research procedures. The interconnection between the three pertinent aspects of research with seven stages of research process is shown in the Figure 1 (Graziano & Raulin's Model of Research Procedure):

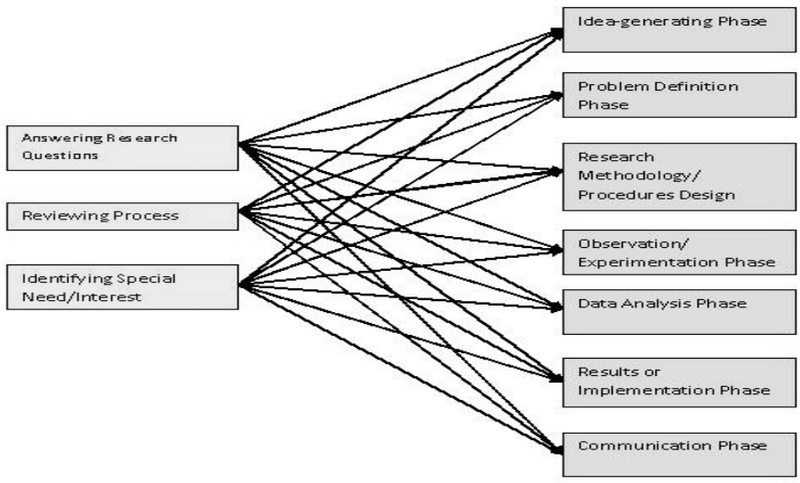

Figure 1 Interconnection between scholarly activity and research procedure

3. Research methodology

With the aim of gauging further understanding on the process of scholarly activities in carrying out research, an interview with research practitioners especially in the ICT area is administered in the Faculty of Computer Science and Information Technology, University of Malaya. The total 15 respondents are interviewed in order to seek the answers for the research question based on supervisory perspective, reviewer or academic articles (journals and proceedings) and special needs or research interests in a specific field of research. The interview is done based on the Kampa (2002) study which attempted to identify some important research questions made by scholars. Additional questions in determining the special interest of researchers in different specialization is posed to the respondent to deduce any patterns that distinguish research approaches between these researchers based on their field of specialization.

The respondents were asked to give answers to the following questions in three usual research related activities:

(1) What are the main questions that you would want your new masters or PhD students to locate so that they will became competent in the new research area?

(2) When reviewing a paper for a journal or conference, what are the questions that you would ask yourself in order for you to understand the paper, either to reject or accept the paper for publication?

(3) What is/are the special needs or research interests that you would want your student to capture in your field of research expertise?

3.1 User group

The study is done within the three divided groups of academicians that based on the years of teaching and research experiences in the field of Computer Science (Artificial Intelligence, Computer Technology, Information System, Software Engineering, Multimedia and Information Science). This study is also done to gauge an understanding on the scholarly activity done by researchers in different stages of research experiences.

3.2 Assumptions

The assumption made in categorizing the respondents based on the number of years, the respondents were involved in the research activity either by being an academic supervisor, funded research or individual interest in a particular research. Table 1 shows the formation structure of the interview questions and answers given by the respondent. The respondents are divided into expert researchers (SR) (> ten years of experiences), intermediate researchers (IR) (>5 to <10 years of experiences) and advance beginner researchers (NR) (< 5 years of experiences). Expert researcher is defined as “someone who displays special skill or knowledge derived from training or experience” (Merriam Webster, 2007). These experts posses Ph.D. degree and hold at least an associate professor post. Advance beginner (researcher) is defined as “one who has coped with enough real situations (research work) to note the recurrent meaningful situational components” or in another word possess experience in research but sometimes facing difficulties to distinguish between what is important and what is not (Daley, 1999). The advanced beginners’ group consists of researchers with master degree and has been involved with supervising at least students in masters level. The group in between advance beginner and expert researcher is considered as intermediate researcher which has the “ability to cope with unpredictable situation” (Daley, 1999) such as managing research project without supervision from senior researcher. The intermediate researchers’ group includes those researchers with Ph.D. degree or have involved in scholarly work for more than five years with majority of them hold the post as senior lecturer. The respondents are further divided into groups based on

specialization of research areas.

Table 1 Classification of respondents and interview questions

Specialization	AI			SE			MM			IS			CT			LIS		
Groups	SR	IR	NR	SR	IR	NR	SR	IR	NR	SR	IR	NR	SR	IR	NR	SR	IR	NR
Questions																		
Research questions?	X			X			X			X			X			X		
Reviewing activity?	X			X			X			X			X			X		
Special interest/needs?	X			X			X			X			X			X		

Notes: AI: Artificial Intelligence; SE: Software Engineering; MM: Multimedia; IS: Information Science; CT: Computer Technology; LIR: Library Information Science.

4. Results

Due to the limitation of space, only the answers given by the expert researchers are presented here. Five senior researchers gave the following responses to the first questions:

(1) What are the specific, up-to-date resources (e.g. thesis, journal and conferences) available in the Information Systems area?

(2) How can the existing research be useful for your own research?

(3) How can the future works discussed in the literature help in narrowing down your research focus?

(4) What are the related papers in this area?

(5) Are there any research groups available (online)?

(6) What are the research clusters?

(7) What are the specific resources (journals) of the AI area?

(8) Are there any centers of excellence available?

(9) What are the problem areas which are lacking in solutions, technically and practically?

(10) Are there any centers of excellence available?

(11) What are the specific online resources available (e.g. ICEE, ACM and Springer)?

(12) What are the issues discussed in the literature?

(13) What are the outstanding problems and possible solutions in this research area?

(14) Can methods in different research areas be used to solve the problems in this research area?

(15) What are the loopholes that can be enhanced in this research area?

(16) What is the research direction in this area?

(17) Is there any duplication of research work? If there is any, how would they differ from one another?

(Notes: Redundant answers are taken out from this list)

One of the expert researchers from MM stated that her approach while dealing with students with research topics and without would greatly differ.

“For the students with a particular research topic, I would first asked him/her to brief me on the subject matters according to the four W (What, Why, When, Where) and one H (How) thing and his/her approach towards the research. For students without research topic, I would ask them about their research interests, advice them to seek ideas from their surrounding, commercial products and identifying research problem in the academic

literature. Of course the time taken for these students to become familiar with the area of research will be longer than those students with a research topic, thus a system that would accelerate the research process would be of great help."

The other senior researcher from AI distinguished between her ways of supervising masters and Ph.D. students:

"For Ph.D. students, I would advice them to look at the journal type of article to strengthen the formation of theoretical framework for the research. However, for master's students, although I would encourage them to look at the journal articles, I would suggest them to look for articles in high impact proceedings such as ACM and IEEE. The emphasize of master's students thesis is on the development kind of research and these resources provide ample solution for their problems".

For the second question, diverse answers are given by the respondents:

(1) Is the content enough to answer the research objectives?

(2) What are the research problems and possible solutions?

(3) Did the paper achieve its objectives?

(4) What are the related research areas?

(5) What is the general view of this paper?

(6) Did the paper discuss enough technical content?

(7) Did the conclusion support the objective and research methodology?

(8) Did the author structure the paper (introduction, methodology, findings, and conclusion) accordingly?

(9) Did the title resemble the content?

(10) Did the abstract covering enough information (the goal of the research, the significance of research and research outcome)?

(11) Did the paper flow accordingly (i.e. follow the abstract)?

(Notes: Redundant answers have been taken out from this list)

One senior researcher from IS elucidated that reviewing process is crucial in academic research activity and gave the following statement:

"For me, reviewing articles for journal or for conference proceeding is not about rejecting or accepting the paper. It is about appreciating one's work and helping them to improve the presentation of the research work done for the benefit of other researchers from all over the world. Of course I will follow the guidelines given by the journal or conference organizer."

For question three, the senior researchers gave the following answers:

(1) How much did the system or application developed solve or meet the users' expectation?

(2) What are the intelligent techniques and features involved?

(3) What are the theoretical frameworks involved?

(4) What is the framework derived from the study?

(5) How can the framework or model be translated into a working prototype?

(6) What are the results from the experimental analysis?

(7) What are the available system level approaches (e.g. suitable algorithm in load distribution, factors that increase network performance) for this research problem?

(8) How can the multimedia object be effectively used for computer aided learning application?

(9) How will the integration of multimedia element add value to the existing computing discipline?

(10) What is the prominent methodology in LIS area?

(11) Can the methodology of other research area fit into LIS area?

The results of the interview reveal several important aspects regarding scholarly activity i.e. research. The scholarly activity such as providing research proposals, writing a research paper and preparation for research presentation requires the three activities identified by Boyer (1990) to be done in accordance with the standard required by the institution. Unfortunately, some of the important information is not known to students except for undergoing the tedious process of information seeking. However, in this study we only consider those answers that can be directly manipulated from the metadata. Answers derived from the respondents that need thorough understanding of the paper such as “Did the paper achieve its objectives? What is the general view of this paper?, Did the paper discuss enough technical content? Did the conclusion support the objective and research methodology?” and so forth are not taken into consideration even though few studies in data mining and text analysis did support this kind of queries (Kontostathis, et al., 2003). Moreover, this study attempts to give as much support as possible to quicken the process of seeking the “right” information, so that the student would become more focused and proficient in the early stages of the studies. The results of the interview are further analyzed in the next subsection.

5. Analysis and discussion

The interview took place in the FCSIT, University of Malaya for the period of four weeks. Table 1 above is populated with the answers received by the respondents. The range of answers given by respondents varies based on their research experiences.

Senior and intermediate researchers exhibit more matured and detailed answers to the questioned post. Majority of the respondent (except for one researcher) emphasized on the importance of their student to locate the relevant literature for the specific research topic. On the other hand, advance beginner (in this study, researchers with less than five years of research experience) interviewed tend to emphasized on the “surface” aspect of the research. The answers given are considered “shallow” as majority of them stressed on the importance relevant literatures yet unable to relate other importance criteria such as current trend in research or what are the “in thing in the market now”. Researches have shown that the performance aspects (of researcher) changes from novice to expert (Dreyfuss ＆ Dreyfuss, 1985). Experts working pattern are said to be shifted from dependence on abstract foundation to solid past experiences, view on situation as “unrelated part” to viewing the “situations as part of a whole” and more involved in performing research work (Daley, 1999). It could seen from this study that experienced researchers are more “engaged” in the research of their students in the sense that they would use their past experience as researchers (and students to other researcher) in helping the students in their research work and able to give more relevant guidelines for the students under their supervision. On the other hand, advanced beginner in this study posse a researcher quality because they do value the importance of literature study and also their abilities to successfully supervise undergraduates in their theses. Thus, this research would not only benefits student (postgraduates and undergraduates) but also novice researchers in supervisory position in enriching their knowledge and skills for future supervision activity.

5.1 Research questions posed to students (researchers in supervisory position)

One of the important advices for students in enrolling in a research-based program is to find a good supervisor. From the discussion with colleagues and researchers’ experiences, a “good” supervisor refers to a

person who can properly lead the students in achieving the ultimate goal of research and finish it with flying colors. However, depending on the experience gained (for instance by duplicating the previous supervisor's way of guiding) by the academician in supervisory position, they would expect that the students "to know what to do" as that is the way it works for them before. Table 2 depicts the analysis of respondents' answer based on research questions.

Table 2 Respondents' answer based on research questions

Pertinent Questions /Respondent	Relevant Literatures	Experts	Trend	Institution	Specific Resources
PR1	x		x		
PR2	x	x	x	x	x
PR3	x	x		x	x
PR4	x		x		
PR5	x		x		
PR6	x		x		
IR1	x		x		
IR2	x				
IR3	x				x
IR4	x		x		x
IR5			x		
IR6	x		x		
JR1	x	x			x
JR2	x				
JR3	x				x
JR4	x				
JR5	x			x	
JR6	x				x

From the interview, the regular research questions posed to students are:

(1) What are the related literatures in this field?

(2) Who are the experts (specific researchers) in this research area?

(3) What is the trend now in this field of research?

(4) Which institution (research group, clusters and centers of excellence) is involved in this type of research?

(5) What are the specific resources (e.g. online databases) for this research?

Other important factors that need to be determined while conducting the research activity such as identifying the problem statement, research methods, loopholes that can be enhanced, research directions, similar systems, framework and models used, etc are features that can be extracted from the relevant literatures. Many researches in the field of data mining have been done to solve the problems in identifying and detecting patterns and relationship from text. However, it is not the intention of this research to tackle this type of support as the focus is to provide the users (in our case the undergraduates and postgraduates students) with necessary support while conducting scholarly activity in terms of searching academic materials to assist them in the early stage of research work by manipulating the metadata and ontology.

The "right" resource will not only assist in writing a good research paper or proposal but also speed up the process for lifelong learning such as understanding the pattern of users' seeking behavior in order to organize

academic resources in such a way that would ease the searching process.

5.2 Reviewing activity

Reviewing research papers are also a part of scholarly activities managed by researchers. The process involved in reviewing the academic articles that are the needs to identify the key research questions, identifying current trends in the research and the novelty and original contribution of the paper to the society at large. Thus, support that could help reviewers to answer this type of research questions regarding the paper reviewing would be of great help. Basically, the common guidelines for reviewing a journal article are as follows:

(1) Content review which includes references to relevant literature, assumptions or speculation made based on existing work, significance of results, originality, comparative study and so forth.

(2) Quality of presentation that includes the organizations of figures, grammar, proper citation and references.

The task for content reviewing could become less complicated with a support environment that could provide researcher with fast access to relevant material and specific places (such as relevant IEEE and ACM online databases of conferences proceedings and journals) to find resources in order to justify the significance and originality of the work reviewed.

5.3 Special needs or interest

Every science discipline has its own special interests which would differentiate one field to the other. This element would be a great help to the novice researchers who are looking at specific requirements needed in their field of research. For instance, Kling and Mckim (2000) deduced from their study that researchers in the field of Information System would be concerned in deciding the type of organizational activities that need to be computerized, the criteria for an effective information system in an organization, information system's evaluation and recognizing how people would use an information system in different organizational environments, i.e. coping with organizational change.

From the results of the interview, six special interests and needs have been identified as depicted in the Table 3 below:

Table 3 Special interests/needs in six fields of computer science and information technology

Field	Special interest/ need
Information science	User acceptance model; Research methodology
Artificial intelligence	Intelligent techniques
System and network technology	Protocol for network communication; Algorithm for encryption and decryption techniques; Network management tools
Library information science	Suitable methodology and models for LIS research
Multimedia	Usability study method; Multimedia object in learning application; Cognitive methods of user interaction with multimedia technology
Software engineering	Architectural design of a system; Available framework in software engineering

The specific research needs of each department could be used as part of a requirement specification or competency question that needs to be answered by the ontology that have been created for the field of CS (Ismail, et al., 2007). Ontology is defined by Grubber (1993) as "a formal, explicit specification of a shared conceptualization". In the other words, ontology is a categorization of all of the concepts in some fields of

knowledge, including the objects and all of the properties, relations, and functions needed to define the objects and specify their actions. In the case of EJUM, users will be provided with sets of relevant terms to assist their information searching process. This specification could be used to assist users in identifying the supporting factors that require their attention related to their fields of research in Computer science and Information Technology.

Figure 2.0 illustrates the proposed architectural design of the semantic support for research activities within the field of Computer Science and Information Technology. The features and characteristics of each component will be further explained below.

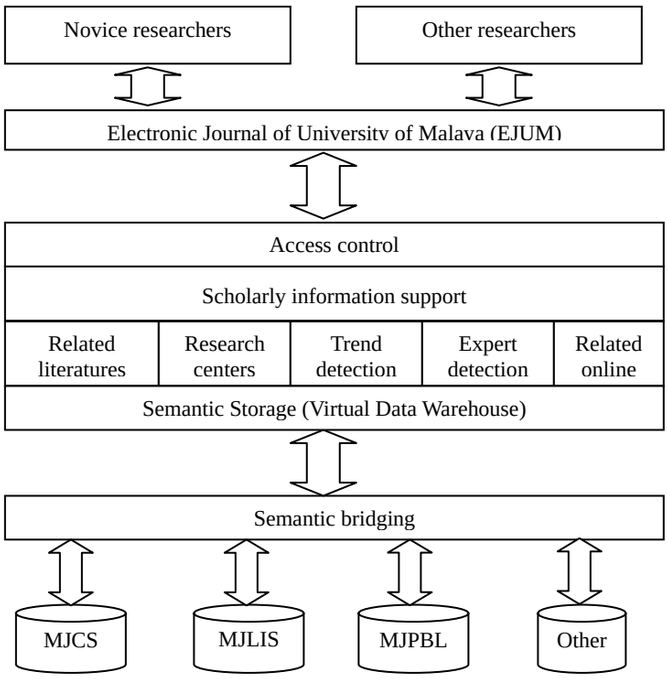

Figure 2 Proposed architectural design of semantic support for research activity

Data of learning resources is gathered through the *Malaysian Journal of Computer Science* (MJCS), *Malaysian Journal of Library & Information Science* (MJLIS) and *Malaysian Journal of Problem Based Learning* (MJPBL) which are being hosted by EJUM (Electronic Journal of University of Malaya). These databases are stored in relational database and for using them semantically, we need a mechanism to map relational data into RDF (Resource Description Framework), because the aim of the Semantic Web (SW) is to give data on the web a well-defined meaning by representing in RDF and relating it to commonly accepted ontologies. D2R Map is a declarative language that used to describe mappings between relational database schemas and OWL ontologies (D2RMap, 2005) in this research. The D2R mappings can be performed by using a D2R processor to export data from a relational database into RDF. After mapping, these RDF files of educational resources will be stored into the Mulgara Semantic Storage that is a database for the storage and retrieval of metadata. Mulgara databases hold metadata in the form of short subject-predicate-object statements, much like the W3C's Resource Description Framework (RDF) standard (Mulgara, 2007). Mulgara uses ITQL (Interactive Tucana Query Language) commands that enable us to query the database and receive results that match the query. ITQL is similar to the

Structured Query Language (SQL) that is used to query relational databases, with some significant differences in syntax and semantic due to the way data is stored in Mulgara.

The scholarly information that supported by this system and will be retrieved through the search function in enhanced EJUM will be the Related Literatures, Research Centers, Trend Detection, Expert Detection and Related Online Resources. Firstly, scholars would select the type of the information they are searching for by selecting the checkboxes related to it. When the query is entered by the scholars through the user interface, the search results will displayed depending on the options selected by the user.

In addition to provide intelligent access to information on EJUM and to find the related literature, ontology can be used for searching enhancement. The Related Literature feature will retrieve information through a matching module which supported by a domain ontology such as ACM Computing Classification System (1998). An ontology based on semantic layer would enable the relationships between the various metadata schemes to be properly represented within the chosen ontology. The ontology could be used as a moderator that facilitates associated searches across heterogeneous data repositories. The ability of using relationship information which is stored in ontology enables semantic search engines to overtake the problems associated with existing search methods. The result of search will be the literatures based on the most related one to the ACM CSS category and 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. Precision and Recall strategy will be used to measure the performance of Information Retrieval (IR) of the system. 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. 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 precision. 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 (Kobayashi and Takeda, 2000) (Table 4 below).

Table 4 The comparison of precision and recall

Precision=	Total number of relevant and retrieved documents by a search engine Total number of results retrieved
Recall =	Total number of relevant and retrieved documents by a search engine Total number of results relevant

Another problem of the generic search engine (as adopted in EJUM) is that the researcher views the paper as an isolated document and they just get the information that is related to the metadata such as author and title. Unfortunately, it fails to associate the results with additional information like other research centers that have similar research. This system is going to provide this ability for the scholars that they can search on the place that authors of the similar literature come from and can make decision such as on the place to further their study in particular field. The problem that we faced to enable this functionality is that the existing EJUM databases has the field which contain the whole address of the institution, and for making this option available we need to have some sort of code manipulation which can just extract the name of the institution to be displayed as the search results.

An emerging trend is a topic area that is growing in interest and utility over time (Kontostathis, et al., 2003). An example of the trend is Semantic Web. The idea of the Semantic Web was appeared in the end of 1999 but it emerged as a trend in the early 2001. Table 5 shows the results of an ACM portal library search on the keyword “Semantic Web” in the title of the documents from 2001 to 2007. Table 5 depicts that research in “Semantic Web” shows and increasing trend from 2001 to 2003. The trend seems to decrease in the year 2004 and 2005, perhaps due to less of interest in such topic. However, the trend seems to emerge in the year 2006 when it was well represented as a topic area in computer science. Emerging Trend Detection (ETD) is the process of finding trends in a specific area in order to provide information on the new developments and research in the area of interest. The researchers need to be informed by the new research activities in their field of the research. In the past they got their required information by reading many documents in their field of research, ask the expert or ones who have read the documents (Kontostathis, et al., 2003). The increasing number of digital repositories results in the growing number of digital information which makes it difficult for scholars to keep up with what is published and developed recently in a specific field by using manual process. This causes the need for an automated trend detection method.

Table 5 Emergence of XML in the 2001

Year	Number of documents
2001	4
2002	26
2003	36
2004	19
2005	15
2006	45
2007	55

In order to detect 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. For instance, counting the frequency of appearance of an expert’s name with respect to certain corpus will yield fast results by just posting a simple SQL in structured database. We will extend the “frequency count method” by learning the nature of name appearance according to the position. For instance, first author will be assigned more weight as compared to second author and so on. A collection of academic articles will be studied to gather information on the maximum number of authors. This will give us a boundary in deducing heuristics for automatically determining an expert in a particular research area (Ismail M. A, 2007).

These are some typically questions that novice researchers have in mind in order to justify other researchers’ level of expertise:

(1) This author has any other papers.

(2) What are the prominent papers in agent systems?

(3) Who are the experts in agent systems?

Resolving those queries requires further investigation and research. Yet, we try to automate this process by suggesting experts’ names to ease the research work.

Another feature that was derived from the interview with the expert is relevant or specific resources of research area. This information is available by doing a thorough analysis on the references part in one research

paper, be it journal, conference proceeding or research report. We will automate this process to give novice researchers some insights in finding the information that they want based on the entered query.

The Access Control Layer (ACL) is the layer where the users need to authenticate themselves before they can use the system. ACL is one part of the system's user interfaces. It will include multiple check boxes which reflect the search branches ("Related Literatures", "Trend Detection" etc.). It also allows user to enter search parameters and retrieve ranked search results.

6. Related work

Few systems are identified to have some similarity with our work. First, an ontology-mediated system named CAS (Computer Science AKTive Space), a portal that allows multiple ways to discover information and rich relations in computer science research (Shadbolt, et al., 2004). The features of CAS are as following:

(1) Browse topics (specific area) and institution for researchers;

(2) Show the geographic range and extent of research topics;

(3) Estimate the experts (in terms of the scholarly impact and cumulative research grant income) in the field of the research by topic and geographic region;

(4) Calculate the researcher's community of practice (the researcher's coauthors and the projects in which they are involved, institutions with which they're affiliated, and topics in which they conduct research);

(5) Identify gaps in research coverage.

CAS can cover the most requirements of the researchers in the field of the computer science but it has some shortcoming, such as it could not define the research trend and the specific databases and online resources for the scholars.

ScholOnto is another ontology-based digital library server for research documents and discourse. It enables researchers to express the additional information for a document and its relationship to the literature through a semantic network (Buckingham, et al., 2000). The features of ScholOnto are as follows:

(1) Enhance individual analysis of literatures for significant conceptual structures;

(2) Provide more effective retrieval of the relevant documents;

(3) Enable structures discourse between researchers (authors can claim their work in this system that are open to contest by others).

ScholOnto has some shortcoming in managing the research-based knowledge. It could not help the scholars to detect the experts of that research, the research trend, the place and institution that the authors come from and the specific databases and online resources for the scholars.

E-Scholar Knowledge Inference Model (ESKIMo) demonstrates the roots of ontological hypertext and scholarly inquiry to provide comprehensive support for research. The core of the ESKIMo system is the ontological model that drives the hypertext (Kampa, 2002). It modeled the research domain in the form of scholarly ontology and used it for discovering information in the web. Our work is much influenced by ESKIMo as it provides the following features:

(1) Project, researchers, and institutes are related to this literature;

(2) Other papers were presented at this conference;

(3) Similar journals in hypertext area;

(4) Experts in knowledge management;

(5) Seminal papers are from the specific computer series of journals;

ESKIMo is limited to hypertext material. It is answers to type of questions posed in the earlier experiment which are directly available in ESKIMo. Our work however, consists of the method to provide homogeneity of heterogeneous resources (different databases in Computer Science and Information Technology field) which is not considered in ESKIMo.

Another project i.e. OntoPortal which is the prequel to ESKIMo have intelligently interlinked hypertext over a research domain, and has been applied to modeling the latest research in metadata. Ontology identifies the concepts such as literature, organisation, standard and their relations within this domain. The OntoPortal system implements this ontological structure over this domain to provide intelligent navigation of the knowledge (Kampa, et al., 2001). The combination of ontological hypertext in OntoPortal and the inference of mechanism enable features such as the following:

(1) Identify the experts in hypertext;

(2) Identify other seminal papers in metadata research;

(3) Detecting the trend of the specific research area;

(4) Providing support for discourse.

OntoPortal overcome the needs of the scholars in the field of their research and include most of the features that they may require. The identification of the research center for the research is one of the requirements that are not considered in this system.

The above systems demonstrate the usage of ontology in providing intelligent inferring for academic-based resources such as proceeding articles and journals. We try to augment our system in order to provide an environment where research process for novice researcher would be ease based on the requirements provided by expert researcher. Even though the solution is far from complete, yet it provides some guidelines for the important features needed on academic-based repository such as EJUM.

7. Conclusion

Based on the interview conducted, five factors have been identified as useful in order to assist users in the early stages of their research activities. These are namely: Related Literatures, Centre of Research Excellence, Research Trend, Experts in Specific Research Area, the Specific Databases and Online Resources.

Knowing these factors earlier would definitely help in accelerating research activities. The results of the interview reveal the true picture related to the problems that students face while dealing with scholarly activity. 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 and from this study, it has been shown that this support is badly in need. Our future work will involve the development of a prototype for incorporating the requirements specified by the researchers in the EJUM environment by utilizing the semantic approach.

References:

Boyer, E. (1990). *Scholarship reconsidered: Priorities of the professoriate*. Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching.

Buckingham, S., Motta, E. & Domingue, J. (2000). ScholOnto: An ontology-based digital library server for research documents and discourse. *International Journal on Digital Libraries*, *3*(3), 237-248.

Daley B. J. (1999). Novice to expert: An exploration of how professionals learn. *Adult Education Quarterly*. Retrieved from http://aeq.sagepub.com/cgi/content/abstract/49/4/133.

D2Rmap. (2005). Retrieved from http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2rmap/D2Rmap.htm.

Dreyfus, H. & Dreyfus, S. (1985). *Mind over machine: The power of human intuition and expertise in the era of the computer*. New York, Free Press.

Graziano, A.M. & Raulin, M.L. (2000). *Research methods: A process of inquiry (4th ed.)*. Allyn and Bacon, Boston.

Finin, T., Ding, L., Joshi, A., Pan, R., Cost, R. S., Peng, Y., Reddivari, P., Doshi, V. C. & Sachs, J. (2004). Swoogle: A search and metadata engine for the Semantic Web. *Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management*, 652-659.

Gruber.T. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, *5*(2), 99-220.

Guha, R., McCool, R. & Miller, E. (2003). Semantic search. *Proceedings of the 12th International Conference on World Wide Web*, Budapest, Hungary: ACM Press.

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.

Ismail M. A., Mashkuri Yaakob & Sameem Abdul Kareem. (2007). Domain specific ontology construction: Falbo's approach. *International Conference on Electrical Engineering and Informatics*, 17th-19th June, 2007, Bandung, Indonesia (ISBN 978-979-16338-0-2), 1051-1054.

Ismail M.A., Mashkuri Yaakob & Sameem Abdul Kareem. (2007). Semantic search engine in institutional repository: An ontological approach. *International Conference on Library, Information and Society 2007*, Kuala Lumpur, Malaysia.

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

Kampa, S., Miles-Board, T., Carr, L. & Hall, W. (2001). *Linking with meaning: Ontological hypertext for scholars (technical report ECSTR-IAM01-005)*. Electronics and Computer Science Department, University of Southampton.

Kling, R. & McKim, G.W. (2000). Not just a matter of time: Field difference and the shaping of electronic media in supporting scientific communication. *Journal of the American Society for Information Science*, *15*(14), 1306-1320.

Kobayashi, M. & Takeda, K. (2000). Information retrieval on the web. *ACM Computing Surveys*, *32*(2), 144-173.

Kontostathis, A., Galitsky, L.M., Pottenger, W.M, Roy, S. & Phelps, D.J. (2003). A survey of emerging trend detection in textual data mining. In: Micheal Berry. (Ed), *A comprehensive survey of text mining*. Springer-Verlag, 2003, 44.

LI X., ZHANG J. & HUANG T. (2007). A standardized learning resources retrieval system based on ontology matching. *Lecture Notes in Computer Science*, *Springer.*

Mulgara - Scalable RDF Database. (2007). Retieved from http://www.mulgara.org/.

Nardi, B. (Ed.). (1995). *Context and consciousness: Activity theory and human computer interaction*. Cambridge: MIT Press, 17-44.

Roy, D., Gevry & Pottenger, W. M. (2002). Methodologies for trend detection in textual data mining. *The Textmine' 02 Workshop*, *Second SIAM International Conference on Data Mining*, 12.

Shadbolt, N., Gibbins, N., Glaser, H., Harris, S. & Schraefelm, M. (2004). CS AKTive Space or how we learned to stop worrying and Love the Semantic Web. *IEEE Intelligent Systems*, *19*(3), 41-47.

Virginia Polytechnic Institute and State University. (2007). Restrieved from http://www.cals.vt.edu/facultystaff/evaluation/scholarlyactivitydefinition.html.

Weiser, C. J. (1996). *The value of a university — Rethinking scholarship*. Oregon State University, Retrieved from http://www.adec.edu/clemson/papers/weiser.html.

(Edited by Lily and Victoria)