An Ontology-based Search Engine for Postgraduate Students Information at the Ministry of Higher Education Portal of Iraq

Hazim N. Abed¹, Alicia Y.C. Tang², Zaihisma C. Cob³, Noor H. Hassoon⁴, and Ahmed L. Ahmed⁵

¹, ⁵University of Diyala, Diyala, Iraq, hazim_numan@yahoo.com
², ³College of Information Technology, Universiti Tenaga Nasional, Kajang, Selangor, Malaysia, aliciat@uniten.edu.my, zaihisma@uniten.edu.my
⁴Noor H. Hassoon, University of Diyala, Diyala, Iraq
⁵Ahmed L. Ahmed, Ministry of Higher Education and Scientific Research, Iraq, Ahmedalany@yahoo.com

Abstract: Information integration plays an important role in the educational environment since it provides a comprehensive view of educational data. The current information portal owned by the Ministry of Higher Education (MOHE) in Iraq often returns the users with a huge load of unorganized information. In order to address this problem we design and develop an ontology-based framework for supporting postgraduate information query at the portal of MOHE. This work uses a methodology adapted from Uschold and King to build the domain Ontology. A query system has been developed to validate the domain ontology for its correctness, usefulness and consistency. The ontology-based query system has also been assessed via questionnaire and hands-on by a group of 34 respondents. Majority of the respondents commented that the query system is able to provide accurate results and that they experienced a more efficient search of information when compared to current system.

Keywords: Ontology; Information integration; framework; Uschold and King; postgraduate query system; protégé.

I. INTRODUCTION

Data integration has emerged as a natural requirement of organizations from the early life of software systems. With the advent of new information technologies such as Internet and intranet, there is an urgent need for information integration from a number of independent and heterogeneous sources. Although organizations such as Higher Education (HE) began to deploy large and complex information systems to support their educational activities, there is still a need for comprehensive information which provides the global view of education process based on different occasions.

HE often needs to combine various types of educational data such as the personal particulars of postgraduate students preferably in a flexible, scalable and personalized manner, in order to provide a global insight into the current state of the individual’s educational processes. Therefore, information integration plays an important role in the educational environment since it provides a comprehensive view of education data [1]. Although databases are employed to represent real world information, they impose several restrictions to the access and management efficient information. The challenge faced by the current ontology engineering processes is dealing with the same kind of problems that databases had to overcome several years ago through a taxonomy [2]. Taxonomy is a hierarchy of categories used to classify documents and other information [3].

However, there is no reliable Web portal that provides structured and accurate information pertaining to the postgraduate studies at the Ministry of Higher Education in Iraq. As a result, this paper attempts to provide solutions to this matter. Hence, this paper is structured as follows: Section 2 provides an overview of information of Iraqi universities and websites. Section 3 presents the overview of using ontology and other related works in construction ontology. Section 4 examines several important aspects of ontology. At the same time, it also presents the procedure and methodology involved in the development of ontology. Section 5 presents the proposed framework, while Section 6 shows the design and development of the postgraduate query system using Protégé platform and query results. Section 7 presents the validation of the query results. Finally, Section 8 provides the conclusion to this research study.

II. IRAQI UNIVERSITIES INFORMATION AND WEBSITE

Iraq has enjoyed a long and proud tradition of distinguished universities; however a sequence of wars and sanctions in recent years have severely damaged the system. After the recent invasion of Iraq, as much as 84% of the infrastructure in institutions of higher learning in Iraq has been burnt, looted or severely destroyed in one way or another. Currently the number of higher education institutions under the management of the Ministry of
Higher Education and Scientific Research are twenty [4]. Each of these universities is equipped with its own website which provide general information about the university including information of the various departments and the activities associated with the university.

The main reason for the establishment of Universities Websites is due to the increasing demand for more reliable and valid websites to represent the image of the Universities as well as to keep up with the rapid pace in this highly advanced technological field where the internet and the websites from an integral and indispensable part in the administrative work of the faculties. Fast communications and speedy responses to different requests have become the actual scale for measuring efficiency in any institutes of learning in the world today. Also the Websites are set up to serve educational purposes such as online publication of lecture materials of the academic staff and the publication of scientific journals through scientific contributions from the postgraduate students of the universities [5].

All Iraqi Universities have set up special units to provide general information about the requirements and mechanisms in the application for postgraduate studies. All these information and materials are mostly in PDF format. The postgraduate units of the different universities are linked to the postgraduate unit of the Research and Development department in the Ministry of Higher Education of Iraq. This unit of Postgraduate Studies in Research and Development department also provides general information, such as the requirements and mechanisms for submission the advertising of postgraduate programs that are available in Iraqi universities. As such, the applicants of postgraduate studies face difficulty in accessing and retrieving the required information. This is because the website normally returns large amount of unstructured data, causing the applicants and users to have difficulty to reading and comprehending the data. In order to overcome this problem a great deal of effort is required to perform data filtration and reorganization before the information can be presented for useful purposes.

### III. RELATED WORKS

The word ontology was taken from Philosophy, where it means a systematic explanation of being. In the last decade, the word ontology began used in artificial intelligence in the 1980s, and is now frequently used by computing and information science audiences [6].

One of the first definitions was given by Neches and colleagues [7], who defined ontology as follows: “an ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary”. This descriptive definition tells what one needs to do in order to build ontology, and gives us some vague guidelines: the definition identifies basic terms and relations between terms, identifies rules to combine the terms, and provides the definitions of such terms and relations. It is noted that, according to Neches’s definition, ontology includes not only the terms that are explicitly defined in it, but also the knowledge that can be inferred from it.

A few years later, Gruber [8] defined ontology as “an explicit specification of a conceptualization”. This definition became the most quoted in literature and by the ontology community. Based on Gruber’s definition, many definitions of what ontology is were proposed. Borst [9] modified slightly Gruber’s definition: “Ontologies are defined as a formal specification of a shared conceptualization”. Gruber’s and Burst’s definitions have been merged and explained by Studer and colleagues [10] as follows: “Conceptualization refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used, and the constraints on their use are explicitly defined. Formal refers to the fact that the ontology should be machine readable. Shared reflects the notion that an ontology captures consensual knowledge, that is, it is not private of some individual, but accepted by a group”.

Many successful stories related to how ontology is used to solve problems in different domains are discussed in the literature. For example, the most recent work in [11] Zemmouchi-Ghomari and Ghomari explained HERO “Higher Education Reference Ontology” ontology building process from requirement specification until ontology evaluation using NeOn methodology. This work can serve as an Instrument for university profiling and strategy development in addition to providing a nondiscriminatory ranking tool. Construction of HERO ontology was undertook by combining three scenario among the nine scenarios proposed by NeOn methodology, namely: development from scratch with reuse of ontological and non-ontological resources in order to achieve a board coverage of relevant concepts describing the knowledge related to the domain of interest.

Laoufi and colleagues [12] presented an architecture of the system MUS which is capable of integrating different ontologies, and moreover to allow reasoning and intelligent information retrieval. They propose an approach based on the e-learning and its components to establish an organizational memory of the scientific, technical and administrative intellectual capital of the university.

In another domain, Brusa and colleagues [13] presented a discussion on the process and product of an experience in developing ontology for the Public Sector whose organization requires a strong knowledge management. Particularly, this process was applied to develop ontology for Budget Domain. They have shown how ontologists could develop domain ontologies by merging different methodologies and software engineering techniques, taking advantages of them. Particularly, this approach has been used to define a Domain Ontology for a Budgetary and Financial System, which could be extended by Task Ontologies and used by different government applications.

### IV. ONTOLOGY DEVELOPMENT METHODOLOGY

Gruber describes ontology as an explicit specification of conceptualization [8]. A conceptualization refers to an abstract with a simplified view of a domain of knowledge (such as medicine, geographic information system, e-government, and education) that is represented for a particular purpose [14]. The domain could be
explicitly and formally represented using existing concepts, objects, entities and relationships that exist between them [15]. Practical descriptions of ontologies have shown their importance in several respects: ontology provides a unifying framework that reduces, and eliminates ambiguities, as well as conceptual and terminological confusion [16]. Also, the ontology can support the sharing and reuse of knowledge. For instance, if a group of researchers wants to create or extend the ontology in a particular field, it can reuse existing ontologies and extend them in order to achieve interoperability and communication among software systems [17]. Although knowledge level modeling area has defined several representation models such as database and ontology, none of them have been popularized as much as ontologies are in the last decade. In fact, ontologies have revolutionized the area of computart science, specifically in artificial intelligence and database disciplines.

Ontologies have become the core components of many large applications. Unfortunately, the training materials are unable to keep up with the growing interest. Thus, it is crucial to design and develop the ontology to meet the demands. There are several steps to adhere in designing and developing an ontology. They are: (1) Determine the domain and scope of the ontology, with consideration of reusing existing ontologies, (2) Enumerate important terms in the ontology, (3) Define the classes and the class hierarchy, (4) Define the properties of class slots, (5) Define the facets of the slots, and (6) Create instances [18].

There are several methodologies available for the building of an ontology. These methodologies may vary in the steps and tasks. The practitioner is provided guidelines on the tasks and steps to perform when building the ontology. The methodology selected for this study is that of Uschold and King. The benefit of this particular methodology is that it is more likely to be understood by novice ontology developers. Also the development of the ontology domain does not take up a lot of time [14]. This methodology provides the guidelines for development of the ontologies include (1) Identify purpose, (2) Building the ontology which is broken down into three steps: Ontology capture, Coding. Integration of existing ontologies, (3) Evaluation, and (4) Documentation. These steps should include a set of techniques, methods, principles and guidelines for each stage [19]. Fig. 1 illustrates the steps and methodology for the development and building of the ontology.

V. FRAMEWORK DEVELOPMENT

Building a framework brings many benefits, for instance findings of the study become comprehensible through it for researchers and practitioners. Additionally, a framework is considered as the infrastructure upon which a study is built, as well as providing the ground for research. The current situation in the department of postgraduate studies in the Ministry of Higher Education in Iraq cannot meet the requirements of the users because the current system does not subscribe to modern techniques and methods in the provision of accurate information for the perusal of the application of a postgraduate program. In this study, the researcher went ahead to present a framework for the process of information integration to support a query system that allows the postgraduate students in Iraq to search for information for their research work. Fig. 2 illustrates the proposed framework.
This work starts with the process of information integration whereby the databases related to postgraduate studies of a few universities were linked to each other. In turn, each university provides this information to the Ministry of Higher Education. As many as four well-known and recognized universities in Iraq have been adopted to perform the process of information integration for this study. The following subsections show the components of the proposed framework.

A. Information Integration Process

Regarding to the information that pertaining the postgraduate students where the Ministry of Higher Education in Iraq is the only party through which the announcement of the study programs available in Iraqi universities as well as requirements for these programs. In fact, each university has postgraduate department content database related to postgraduate students, which functionally linked to research and development department in the Ministry of Higher Education and Scientific Research, through the department of postgraduate studies, for organizing and directing all about postgraduate of orders and regulations and instructions between the Ministry of Higher Education and the University.

Therefore, Ministry of Higher Education has become necessary to use modern techniques to overcome the problems that appear through the use of traditional method of dealing with this information. The process of information integration starts from the Ministry of higher education whereby they need to merges the databases related to postgraduate studies of universities to create a unified database. The main concepts that are captured from the unified database is used to start the design of the ontology. Then, a unified Ontology is produced to gather the information from different sources.

B. Ontology Construction

After completing the process of information integration in the Research and Development Department in the Ministry of Higher Education in Iraq, and get a unified database contain all information pertaining to postgraduate studies in Iraqi universities. The process of design and develop ontology will start using the information in the unified database in order to construction a query system able to retrieval faster and structured information rather than the current system adopted by the MOHE.

Ontology repositories is an important component in the ontology construction which can help users search and use the ontologies on the Semantic Web. The ontology repositories are used for purposes of publication, sharing, and the reuse of ontologies including vocabularies for content indexing, information retrieval, and content integration [20]. Ontology repositories are needed to store and maintain ontologies.

C. SPARQL Processes

SPARQL (Standard Protocol for RDF Query language) is a program that is assigned the task to extract information from RDF graphs for machine understandable representation. SPARQL is programmed for a specific purpose, for instance, transferring meaningful information from one machine to another. Another example is the use of semantic search engines and browsers for semantic traversal [7]. SPARQL supports a number of tasks including aggregation, sub-queries, negation, and creation of values by expressions, extensible value testing, and constraining queries by the source RDF graph. All the tasks use SPARQL for making enquires of the domain ontology. The results of SPARQL queries can be found in the form of RDF graphs [21].

D. Results of Query and Assessment of the Results

The user’s keywords that they would like to get answers that are related to these keywords. In this research study, Queries Tab in protégé is used to help the users to enter their query and obtain the result. The results will be evaluated through presented them to the users.

VI. DESIGN AND DEVELOPMENT OF THE POSTGRADUATE QUERY SYSTEM

The methodology used in this work are adapted from Uschold and King Methodology and described in section IV. The design of the ontology component describes the construction of ontology which forms the base for knowledge of the postgraduate query system.

The first step entails the proposal of informal competency questions which help to identify the scope of the postgraduate query system. The competency questions are important as they help to build the ontology that is accurate and provides precise information to the user. Competency questions were designed to collect data for further requirements [20]. The competency questions were proposed based on informal interviews conducted with many of the lecturers, postgraduate students and administration staff of postgraduate Studies unit at one of universities mentioned in this study. Since the ontology is used in a query system, such questions can help achieve accuracy in the design of the ontology. The ontology must be able to represent the questions using its terminology and the answers are based on the axioms. Table 1 illustrates the list of competency questions supported by the ontology.

<table>
<thead>
<tr>
<th>Q</th>
<th>Competency questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Who is the expert in a particular research area and in which university is he/she working at? What are the names of students who are under his/her supervision?</td>
</tr>
<tr>
<td>2</td>
<td>What are the certificates required for the application of postgraduate studies?</td>
</tr>
<tr>
<td>3</td>
<td>What is the name, gender and qualifications of a particular postgraduate lecturer at the certain university? What are the names of each postgraduate student under his/her...</td>
</tr>
</tbody>
</table>
supervisor?

4. What are the entry requirements for a particular program?

The second step requires the listing of all the concepts used in the constructing ontology as well as their relationships in the chosen domain. Table 2 illustrates the main “Concepts” used in the domain that are identified in order to construct the ontology.

**Table 2: Main concepts used in the domain**

<table>
<thead>
<tr>
<th>Class</th>
<th>Subclass</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>Diploma Student</td>
<td>This class contains all the essential data about personal particular of the postgraduate students such as name, ID, research mode, the program of study, name of the supervisor and name of university he/she is studying in.</td>
</tr>
<tr>
<td></td>
<td>MSc Student</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PhD Student</td>
<td></td>
</tr>
<tr>
<td>Lecturer</td>
<td>Nil</td>
<td>This class presents the profile of the postgraduate teaching staff such as lecturer's name, area of expertise, name of university, he/she is presently employed, gender and academic qualifications.</td>
</tr>
<tr>
<td>University</td>
<td>University Name</td>
<td>This class presents information of the Iraqi universities that include name, location and the rank of the university and the programs offered</td>
</tr>
<tr>
<td></td>
<td>University Rank</td>
<td></td>
</tr>
<tr>
<td></td>
<td>University Location</td>
<td></td>
</tr>
<tr>
<td>Application Process</td>
<td>Admission Process</td>
<td>This class informs the applicant of the processes to be adhered to during the application for a particular program</td>
</tr>
<tr>
<td></td>
<td>Document Process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Selection Process</td>
<td></td>
</tr>
<tr>
<td>Entry</td>
<td>Higher</td>
<td>This class informs</td>
</tr>
</tbody>
</table>

Finally, the ontology is constructed using the Ontology editing tool known as Protégé. The protégé platform supports two main ways of modeling ontologies through the Protégé-Frames and Protégé-OWL editors [22]. Fig. 3 illustrates the visualization of ontology using OWLVIZ tab in protégé. OWLViz is one of official protégé plugins that permits visualization and navigation of the class hierarchy of an OWL ontology. It allows the asserted and inferred classification hierarchies to be visualized [23].
Figure 3: The ontology for Postgraduate Query System visualized in OWLVIZ

Figure 4: Results of query using the competency

(a) Results

(b) Results

(c) Results

(d) Results
Fig. 4 shows the query results based on the questions presented in Table 1, using open SPARQL query panel.

Fig. 4a shows the name, and academic qualifications of lecturers who are experts in the field of electrical engineering and the students under his/her supervision as well as the names of the university where he/she is attached. The information presented in Figure 4a is useful for applicants to gather information about the universities which offer the program they are interested in and the lecturers who are experts in that particular program as well as the names of the students who are currently studying in this program. For the applicants, acquiring such precise information is vital in selecting the best-fit postgraduate programs.

The results tabulated in Fig. 4b presents all the certificates required for application to postgraduate program. This information informs the applicant of the exact certificates required for a particular program that the applicant wants to pursue in. Fig. 4b shows a list of certificates required for a Master degree in IT. The applicant should possess a Graduate Certificate in the relevant program, a certificate in Computer Skills, TOEFL ITP Certificate and a certificate of No Objection Staff Certificate.

Fig. 4c tabulates the biodata of all the postgraduate lecturers at Diyala University, together with the names of students who are under their supervision. Figure 4c displays the accurate information of the teaching staff attached to the Department of postgraduate studies at Diyala University together with the names of students who are under their supervision. This information proves to be very useful for postgraduate students who have to decide on the right place and a suitable supervisor for the program of study.

Fig. 4d provides a list of all the requirements that are required of a candidate in the application for a Master program in Dental.

VII. VALIDATION OF THE QUERY SYSTEM

The testing of ontology-based query system was conducted to identify the advantages of the system over the existing ministry of higher education portal of Iraq. To do this, the researcher must analyze the current system and identify the differences between the current system and the new one as well as to prove that it behaves and produces the results expected by the users. The respondents of this questionnaire consist of postgraduate students and lecturers of University of Baghdad. The present system is presented to the users for evaluation. Then the users answer questions constructed to evaluate the acceptance of the new system. The questions are adapted based on [24] that conducted the validation of a questionnaire on the user satisfaction of a computer user to measure information system success in small business.

The research sample was randomly selected. The number of respondents selected for this research is 34. From this sample, there were 2 Higher Diploma students, 8 MSc students, 19 PhD students, and 5 lecturers who answered the questionnaire. All the responses to the test questions and the data obtained from these questions were processed using the SPSS Package. From the data collected, the Mean was calculated by the SPSS. The Mean value ranged from one to five. All these items were placed into a Likert scale ranging from 1 to 5, where 5 -Strongly Agree (SA), 4 - Agree (A), 3 - Neutral (N), 2 - Disagree (DA), and 1- Strongly Disagree (SDA). Table 3 shows the results of the questions related to the evaluation of ontology-based query system.

Table 3: Results of the evaluation of questions on the query system

<table>
<thead>
<tr>
<th>Item Description</th>
<th>SA</th>
<th>A</th>
<th>N</th>
<th>DA</th>
<th>SDA</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>The query system provides me with sufficient information.</td>
<td>13</td>
<td>18</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>4.26</td>
</tr>
<tr>
<td>The query system provides reports that seem to be just about exactly what I need.</td>
<td>9</td>
<td>19</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>4.06</td>
</tr>
<tr>
<td>The query system is accurate.</td>
<td>22</td>
<td>7</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>4.50</td>
</tr>
<tr>
<td>The query system provides me with up to date information.</td>
<td>6</td>
<td>17</td>
<td>9</td>
<td>2</td>
<td>0</td>
<td>3.79</td>
</tr>
<tr>
<td>The information I get from the query system is clear.</td>
<td>8</td>
<td>19</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>4.03</td>
</tr>
<tr>
<td>Using the query system saves me time.</td>
<td>19</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.56</td>
</tr>
<tr>
<td>Using the query system improves job performance.</td>
<td>4</td>
<td>19</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>3.76</td>
</tr>
<tr>
<td>The query system is easy to learn.</td>
<td>11</td>
<td>17</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>4.06</td>
</tr>
<tr>
<td>Query system is easy to use.</td>
<td>10</td>
<td>19</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>4.12</td>
</tr>
<tr>
<td>The query system is complex.</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>14</td>
<td>11</td>
<td>2.00</td>
</tr>
</tbody>
</table>

From the findings, it is clear that the respondents are satisfied with this query system in the search for information. Majority of the respondents commented that the query system is able to provide accurate results, and that they experienced a more efficient search of information when compared to current system in the department of postgraduate studies in the Ministry of Higher Education in Iraq which cannot meet the requirements of the users because the current system does not subscribe to modern techniques in providing fast and accurate information for applicants of postgraduate programs.

VIII. CONCLUSION

This paper presents a framework for the process of information integration to support a query system that allows the postgraduate students in Iraq to search for information for
their research work. The design and development of the domain ontology is based on the Uschold and King methodology, while the ontology is built using the Protégé ontology modeling tool. The validation and evaluation were conducted to make sure that all the components of the framework are integrated and arranged in a sequential manner as well as to ensure the system is accepted for its accuracy and distinct features. The results of the queries showed that ontology-based system is able to assist postgraduate students locate precise, and relevant information for postgraduate studies at the universities. It is hoped that this study will promote the adoption of semantic technologies in academic administrative processes in the Ministry of Higher Education in Iraq.

REFERENCES


Authors Biographies

Hazim N. Abed was born on the 12 of August, 1985 in Diyala state, Iraq. He received his Bachelors in Computer Science from Diyala University, Iraq in 2007, and Master in Information Technology from Universiti Tenga Nasional (UNITEN). His major filed of research are semantic web, Ontology, Knowledge representation, and cloud computing. He is currently teaching at Diyala University, College of science, and Department of computer science

Zaïhisma C. Cob is a Lecturer at the Information Systems Department, College of Information Technology (COIT), UNITEN. She obtained her Bachelor in Information Technology from Multimedia University, Cyberjaya and Master in Management (Information Technology) from Universiti Putra Malaysia. Her research interests include
Knowledge Management and Semantic Web. Currently, Zaihisma works on applying data mining to business applications, ontology-based KMS and ontology development to Semantic Web applications. She is a member of IEEE, ACM Computer Society and the Knowledge Management Research Group at COIT.

Alicia Y.C. Tang is a Senior Lecturer at the Systems and Networking Department, College of Information Technology (COIT), UNITEN. She received her Master of Science from University of Manchester, and her Ph.D. in Artificial Intelligence from University of Malaya. Her research interests include qualitative reasoning, data mining, ontology, AI in education, and agent technology. Currently, Dr. Alicia works on applying data mining to business applications, agent technology to Learning Management Systems (LMS), and ontology development to Semantic Web applications. She is a member of the Asia Pacific Society of Computer in Education, IEEE, AIED Society, and the Centre for Agent Technology at COIT. Alicia has served as a reviewer for a number of International Conferences and Journals, and has been member of the program committees for several workshops and conferences.

Noor H. Hassoon was born on the 29th of December, 1985 in Diyala state, Iraq. She received her Bachelors in Computer Science from Diyala University, Iraq in 200, and master in Information Technology form Universiti Tenaga Nasional. Her research interests include LMS, Cloud Computing and AI in education. she is currently teaching at Department of computer science , College of Education for Pure Science, Diyala University.

Ahmed L. Ahmed was born on the 27th of January, 1980 in Al-Anbar state, Iraq. He received his Bachelors in Computer Science from Baghdad University, Iraq. He graduated 2005.and Master in Information Technology from Universiti Tenaga National. His research interests in cloud computing technology, Learning Management System, Software Agent. He is currently working at Ministry of Higher education in Iraq.