1-1-2013

Ontology-based search system using hierarchical structure design

Norasykin Mohd Zaid
Universiti Teknologi Malaysia, nmz056@uowmail.edu.au

Sim Kim Lau
University of Wollongong, simlau@uow.edu.au

Mohd Nihra Haruzuan Mohamad Said
Universiti Teknologi Malaysia

Ontology-based search system using hierarchical structure design

Abstract

Major retrieval issues on conventional online database search systems faced by novice researchers are identified and described. An ontology-based search framework is proposed. The framework consists of five elements: (1) user, (2) query, (3) query interface, (4) search process, and (5) result. The query interface is designed based on the ontology hierarchical structure in the ontology datastore and is represented as keywords/sub-keywords structures in the query interface. Ontology enables relationships between keywords and terms to be defined. Ontology allows desired information to be retrieved by sharing common vocabularies with an understanding of meaning of terms in the domain. A prototype system is developed based on the design science methodology selected to guide the system development. The system will process the search query once the query is sent. The search process will look for the matched documents and return results to the user. The results that are returned to the users will be reviewed and users have the option to repeat the process if the results do not fulfil the query made. The system development takes into account problems identification in the database search environment and motivation in improving and extending current limitations.

Keywords

structure, search, system, design, hierarchical, ontology

Disciplines

Engineering | Science and Technology Studies

Publication Details

Mohd Zaid, N., Lau, S. Kim. & Mohamad Said, M. (2013). Ontology-based search system using hierarchical structure design. In C. Teh, H. Chae, S. Zaidi Adruce, P. Anding, C. Chen, N. Aziz & K. Tan (Eds.), The 9th International Conference on Cognitive Science (ICCS 2013) (pp. 566-570). Netherlands: Elsevier BV.

Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 97 (2013) 566 – 570

Procedia
Social and Behavioral Sciences

The 9th International Conference on Cognitive Science

Ontology-based search system using hierarchical structure design

Norasykin Mohd Zaid[a,*], Sim Kim Lau[b], Mohd Nihra Haruzuan Mohamad Said[a]

[a]*Universiti Teknologi Malaysia, UTM Skudai, 81310, Malaysia*
[b]*University of Wollongong, Wollongong, 2500, Australia*

Abstract

Major retrieval issues on conventional online database search systems faced by novice researchers are identified and described. An ontology-based search framework is proposed. The framework consists of five elements: (1) user, (2) query, (3) query interface, (4) search process, and (5) result. The query interface is designed based on the ontology hierarchical structure in the ontology datastore and is represented as keywords/sub-keywords structures in the query interface. Ontology enables relationships between keywords and terms to be defined. Ontology allows desired information to be retrieved by sharing common vocabularies with an understanding of meaning of terms in the domain. A prototype system is developed based on the design science methodology selected to guide the system development. The system will process the search query once the query is sent. The search process will look for the matched documents and return results to the user. The results that are returned to the users will be reviewed and users have the option to repeat the process if the results do not fulfil the query made. The system development takes into account problems identification in the database search environment and motivation in improving and extending current limitations.

Selection and/or peer-review under responsibility of the Universiti Malaysia Sarawak.

Keywords: ontology search; ontology datastore; novice researcher.

1. Introduction

Online database systems have been used widely by researchers when conducting research for research topics. However, to be able to process queries efficiently is not an easy task. Common problems encountered in full text searches include lack of precision and relevance. [1] explains the weakness of full text searches and can include synonym and homonym, variant spellings, short form of terms and different languages or dialects used. These problems can be categorised into five groups: (1) query formulation, (2) lack of user perseverance, (3) terminology and search terms, (4) lack of user experience, and (5) synonyms and homonyms.

The concept of information retrieval was introduced when we began to store information in computerised databases. An information retrieval technique is required to retrieve the required records when needs arise. Research has been conducted to improve capabilities and facilities in the area of information search and retrieval [2, 3, 4]. The aim of the information retrieval process is to help users to search and use information efficiently [5]. Large *et al.* [6] describes the aim of information retrieval is to match information needs with items in the database. Extensive research has been conducted in the field of information retrieval to assist people in information seeking and various

* Corresponding author. Tel.: +6-010-406-1060.
E-mail address: norasykin@utm.my

1877-0428
Selection and/or peer-review under responsibility of the Universiti Malaysia Sarawak.
doi:10.1016/j.sbspro.2013.10.274

definitions of information retrieval have been defined [5, 7, 8, 9, 10, 11, 12, 13, 14]. [15] describe information retrieval as a process of how information is being stored, organised, represented and easily accessed by the user. While [16] describes information retrieval as retrieving documents to meet the user's scope of search. [17] define information retrieval as a specific condition that concerns searching and retrieving of information stored in a database.

Issues of database searching and retieval methods have been presented in this paper. The issues identified show that current approaches of database search design requires users to have a certain level of skill in search techniques, such as knowing how to use appropriate keywords and Boolean operators to search. Novice researchers often have difficulties in conducting searches using traditional database systems due to a lack of search skills. The novice researchers find it difficult to formulate a query. They find it difficult to transfer what they have in mind to correct search query formats. In addition, the search results can sometimes be unmatched to what users have expected, retrieving too many results or nothing at all. This problem is compounded with terminology problems as some users find it difficult to use the same keywords as what is indexed in the database. They also face problems when they need to use keywords that are different from their native language. Very often, the system cannot conduct bilingual searches. Some systems use bilingual searches consisting of cross-language and multi-language information retrieval to overcome the problem. Major problems related to search queries can be categorised into: (1) query formulation, (2) lack of user perseverance, (3) terminology and search terms, (4) lack of experience in searching, and (5) synonyms and homonyms.

Recently, Semantic Web technology has been used in information retrieval. In this chapter, issues relating to Semantic Web architecture as well as the capability of Semantic Web technology in understanding the meaning of information across the Web have been presented. Ontology which is used to represent a domain of knowledge to allow use, reuse and sharing have also been described. Knowledge sharing through ontology data-store can be achieved by creating the ontology either from scratch, mapping into similar domains or merging between related domains. Ontology can be used to define the meaning of data and relationships between resources. An ontology-based search approach can provide an opportunity to overcome problems related to search queries.

2. Ontology-based Search Approach

Problems associated with information retrieval as described above can be overcome by using an ontology-based search approach. The use of ontology can help to formulate a query, which can be organised by defining a relationship of concepts using a hierarchical structure in the form of class/subclass relationship. Ontology for a domain of specific content can be developed to be used to formulate a query. The ontology datastore can also be restructured easily when a relationship of concepts needs to be updated. With the proposed ontology approach, database content can be referred to by defining properties and relationships for instances/individuals in the datastore.

3. Evaluation

The purpose of the evaluation process is to enable current students enrolled in the Education Faculty, Universiti Teknologi Malaysia, to evaluate the prototype system. The respondents are Year 1 students who have enrolled in the degree programmes at the Faculty of Education, Universiti Teknologi Malaysia. These respondents are novice researchers who have only basic research skills and are inexperienced in academic research.

Due to the bilingual nature of the students enrolled in the Faculty, the questionnaires were distributed in two languages: English and Bahasa Malaysia. However, respondents have chosen to complete the questionnaire in the Bahasa Malaysia language. The evaluation of the prototype was conducted in the computer laboratory. The survey was supervised by a lecturer and one technical support staff member was assigned in each laboratory (three computer laboratories were used during the survey).

3.1.1. Evaluation based on Student Survey

To complete the survey students are asked to access to the server using the link provided to use the prototype. The server is provided by the Faculty of Education. The system has been uploaded remotely using an ftp server which is WinSCP version 4.3.5. The prototype system has been uploaded to the server a

month before the survey to allow the system to be tested and be functional before the student evaluation is conducted.

In this survey, students have to answer 18 Likert scale questionnaires. The Likert scale is made up of 1 to 5 scales. 1 refers to "strongly disagree", 2 refers to "disagree", 3 refers to "slightly agree", 4 refers to "agree" and 5 refers to "strongly agree". These questions aim to evaluate the student's experience in using the prototype system. The questions relate to the ability of the system to facilitate the search process and the user friendliness of the system.

Table 1: Results of questionnaire

Number	System Evaluation	Mean	Standard deviation
	Part 1: System usage		
1	The system gives me greater control to find what I need	3.47	0.85
2	The system enables me to search the thesis quickly	3.62	0.90
3	The system returns the search results more accurately	3.57	0.89
4	The system returns the search results containing theses written in English and Bahasa Malaysia	3.84	0.78
5	The drop-down menu makes the searching process easier	3.75	0.92
6	The drop-down menu gives me limited searching to the thesis database	2.56	0.99
7	I need to spend a lot of time to learn how to use the system	3.45	1.11
8	I prefer to use the drop-down menu rather than an open keyword search when I am not sure what topic I am looking for	3.48	1.03
9	I feel very confident in using the system	3.48	0.93
	Part 2: System interface		
10	When I start to use the system I know where to start the searching process	3.04	0.99
11	When I use the system I know how to do the searching	3.22	1.01
12	I can easily follow the search steps to get to the search results	3.41	0.80
13	Using the system requires a lot of mental effort	3.60	0.79
14	I like the layout of the system	3.56	0.88
15	The system navigation features are difficult to use (buttons, scrolls, etc)	2.85	0.93
16	The system provides helpful guidance in performing tasks	3.45	0.79
17	Overall, I found that this system is useful in getting the thesis as needed	3.80	0.92

The evaluation results in questionnaire section, shows that the majority of respondents expressed a positive viewpoint on the use of drop-down menus to make the search process easier and quicker, and they do not need to consider which keyword to use. Evaluation results also show that the prototype system is able to return search results in both English and Bahasa Malaysia language. Generally, most of the respondents agreed that the system is useful to obtain the thesis title they need and they can follow the search sequence in achieving the results.

3.1.2. Evaluation based on search experiments (using a prototype)

Experiments have been conducted to evaluate search outcomes using the proposed ontology-based prototype system and the database system, which is non-ontology based. The experiment was conducted using scenarios given

to the respondents. Researcher founds that the search result returned from the non-ontology based can lead to unsuccessful search rather than ontology-based prototype system that can conduct a simple search as well as a complicated search. Table 2 shows the comparison of search results

Table 2. Comparison of the ontology-based system prototype and the current database system

Scenario	Results based on exact phrase using current database system	Results based on all words using current database system	Results based on any word using current database system	Results obtained from proposed ontology-based prototype system
Scenario 1: *"Lecturers' and TESL teacher trainees' perceptions towards the level of readiness in performing teaching practice".*	1 record found.	1 record found.	8103 records found.	5 records found.
Scenario 2: *"You are required to do research for your course presentation. You have been assigned a topic related to 'teaching and learning in Mathematics'. You have to use the database system provided to search for those related theses and present the outcome about the choice of topics that have been done by the previous students".*	No record found.	No record found.	No record found. However, when searching the topic separately; 939 records found for 'teaching mathematics' and 1005 records found for 'learning mathematics'.	11 records found.
Scenario 3: *"An investigation on approaches used to teach literature in the ESL classroom: A case study of Sekolah Menengah Kebangsaan Taman Desa Skudai, Johor Bahru".*	No record found.	1 record found. (Exact title included)	Microsoft JET Database Engine error '80004005' Query is too complex.	3 records found.

The results have shown that the ontology-based prototype system is able to perform a more effective search compared to the existing database system, which is non-ontology-based.

4. Conclusion

This research describes the application of ontology-based search framework to address information retrieval problems faced by novice researchers. The six steps of design science research process introduced by Ken *et. al.* [18] have been adapted. The six steps are: (1) problem identification and motivation, (2) objectives of solution, (3) design and development, (4) demonstration, (5) evaluation, and (6) communication. The unique query interface is formulated as ontology using a mind-map to enable a dynamic query to be displayed in a hierarchical structure. The evaluation of the case study has been conducted using three case scenarios.

The results show that respondents found that the use of ontology as a query interface has helped in a query formulation. Results also showed that 60.10% of respondents were able to use keywords for their first choice to make retrieval. The majority of respondents agreed that they could use the prototype system confidently to conduct a search query and the prototype system is useful in finding thesis resources.

Acknowledgements

The authors would like to thank Universiti Teknologi Malaysia, Ministry of Education Malaysia (MoEM) and University of Wollongong, Australia for their support in making this research possible. This research was supported by Research University Grant (Q.J130000.2431.00G83) initiated by UTM and MoEM.

References

[1] Beall, J., The Weaknesses of Full-Text Searching. *The Journal of Academic Librarianship*, 2008. 34(5): p. 438-444.

[2] Bates, M.J., The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review*, 1989. 13(5): p. 17.

[3] Klink, S., et al., Browsing and Visualizing Digital Bibliographic Data, in *Joint EUROGRAPHICS - IEEE TCVG Symposium on Visualization (2004)*, C.H. O. Deussen, D.A. Keim, D. Saupe, Editor. 2004.

[4] Malki, M., et al., Towards ontology extraction from data-intensive web sites: An HTML forms-based reverse engineering approach. *International Arab Journal of Information Technology*, 2008. 5(1): p. 34-44.

[5] Becker, J. and R.M. Hayes, *Information storage and retrieval: tools, elements, theories*. 1963, New York :: Wiley. 448.

[6] Large, A., L.A. Tedd, and R.J. Hartley, *Information seeking in the online age: principles and practice*. 1999, London ; New Providence, NJ :: Bowker-Saur.

[7] Hancock-Beaulieu, M., Evaluating The Impact Of An Online Library Catalogue On Subject Searching Behaviour At The Catalogue And The Shelves. *Journal of Documentation*, 1990. Vol. 46(Iss: 4): p. pp: 318-338

[8] Hancock-Beaulieu, M., A comparative transaction log analysis of browsing and search formulation in online catalogues. *Program: electronic library and information systems*, 1993. Vol. 27(Iss: 3): p. pp.269 - 280.

[9] Korfhage, R.R., *Information storage and retrieval*. 1997, New York:: Wiley Computer Publishing.

[10] Jansen, B.J. and U. Pooch, Web user studies: A review and framework for future work. *Journal of the American Society of Information Science and Technology*, 2001. 52(3): p. 235 - 246.

[11] Jansen, B.J. and A. Spink, How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management*, 2004. 42(1): p. 248-263.

[12] Marchionini, G., S. McDonald, and J. Tait, From Information Retrieval to Information Interaction, in *Advances in Information Retrieval*. 2004, Springer Berlin / Heidelberg. p. 1-11.

[13] Marchionini, G. and R. White, Find What You Need, Understand What You Find. *International Journal of Human-Computer Interaction*, 2007. 23(3): p. 205-237.

[14] Chowdhury, S., F. Gibb, and M. Landoni, Uncertainty in information seeking and retrieval: A study in an academic environment. *Information Processing & Management*, 2010. 47(2): p. 157-175.

[15] Baeza-Yates, R. and B. Ribeiro-Neto, *Modern information retrieval*. 1999, New York: Harlow, England:: ACM Press ; Addison-Wesley. 513.

[16] Chowdhury, G.G., *Introduction to modern information retrieval*. 2nd ed ed. 2004, London:: Facet Publishing. 474.

[17] Meadow, C.T., et al., *Text information retrieval systems*. 3rd ed. ed. 2007, Amsterdam; London:: Academic Press.

[18] Ken, P., et al., A Design Science Research Methodology for Information Systems Research. *J. Manage. Inf. Syst.*, 2007. 24(3): p. 45-77.