Ontology Based Scientific Keywords Recommendation System under Web 2.0

Na Xue, Su ling Jia, Jin xing Hao and Qiang Wang
Beihang University, Beijing, China

Abstract—As the born and communication base of new science and new technology, scientific research requires effective information integration and knowledge management to improve the efficiency of scientific research. The growth of e-science in circumstance of Web 2.0 has created a need to integrate large quantities of diverse and heterogeneous data. To tackle with the scientific information management problems, we proposed an ontology-based scientific keywords recommendation system under web 2.0. The main goal achieved is to extract the meaningful information and recommend to web scholars through the proposed system. The components of the system are Integration Interface, Service Module, Text Processor, Recommendation Module and Ontology Database. Experimental results and performance evaluation shows that the proposed system provides the effective way to recommend semantic related keywords for scholars.

Index Terms—Keywords recommendation, Ontology, Recommendation system, Scientific information, Web 2.0.

I. INTRODUCTION

Web 2.0, the second generation of the Internet, allows people to produce and debate information online [1]. The Internet has evolved from a "Web" where information could be browsed and sought within a vast network of pages into a collection of social applications where individuals contribute, aggregate, "tag", and exchange materials including messages, votes, and other social information [2]. It provides the technical platform for wide communication and cooperation. Users are engaged more actively in the exchange and construction of web content than ever before. Therefore, the role of the web in our daily lives has been much changed. Instead of passively accepting the materials, online users begin to actively create webpage messages. With their own contribution to the web, the interaction between users and web content becomes more quickly and effective. This transformation is "web widely", which is certainly world widely. And there is no exception for the academia.

As the born and communication base of new science and new technology, scientific research requires effective information integration and knowledge management to improve the efficiency of scientific research. The growth of e-science in circumstance of Web 2.0 has created a need to integrate large quantities of diverse and heterogeneous data. Coupling with the scientific community, Web 2.0 brings about the open science reformation in the internet and information age. The Open

Science (or Open Research, Open Source Science, Science 2.0) becomes research hot spot problems. This inevitably causes exponentially growth of the online scientific information. For the precious time of scientific research, scholars need to seek out the closely correlative information.

Information retrieval through keywords is an efficient way to find the required information. Similarly, scholars are searching scientific literature through keywords to filter the materials. Many specialty domestic and external websites adopt the keyword-based query method. For example, there are CNKI in China and IEEE Xplore in USA. However, keyword-based searching engines have several problems. Firstly, the keywords entered by users were incomplete and imprecision, which could not clearly indicate the query demands of users. A small part of the return results are required. Therefore, more time still to be taken to filter the results and start further searches. Secondly, traditional keyword-based methods often ignore the semantic meaning of words and phrases. In other words, many keywords that are the same words have different meaning in different fields or contexts. And in different literatures, the same keywords are treated as different things from specific perspectives by different authors. There also have different expressions to represent the identical meaning. For example, search with the keyword "ontology". The results are interfused with domain of philosophy and artificial intelligence. These conditions multiply the difficulties to filter interesting literature for each scholar. Therefore, the keywords determination and searching method need to be improved. Typically, keywords selection should include semantic relations.

According to the above description, the following problems need to be figured out:

(1) How do scholars know the close related keywords they are searching for?

(2) In which form the keywords should be organized to share with peers?

(3) How to interpret the different understanding of keywords from different authors in Web 2.0 circumstance?

(4) What is the effective mechanism should be to guarantee achieving the common cognition?

All the above described questions to the scientific information management problem should be answered by further improving the content classification and index of web resources. To tackle with these problems, we proposed an ontology-based scientific keywords recommendation system under web 2.0.

Ontology in philosophy refers to a conception of what can exist or 'be' in the world. In artificial intelligence, it was introduced to represent the construction of knowledge models, which specify concepts or objects, their attributes, and interrelationships. A knowledge model is a specification of a domain, or problem-solving behavior, which abstracts from implementation-specific considerations and focuses instead on the concepts, relations, and reasoning steps characterizing the phenomenon under investigation [3]. Ontology has become the knowledge representation medium of choice in recent years for a range of computer science specialities including the Semantic Web, Agents, and Bio-informatics [4]. It can be used for conventions of the knowledge-level communication: content-specific specifications [5]. Common ontology plays a central role in achieving sharable and reusable knowledge bases.

The application of ontology and knowledge modeling in our study is to specify and implement a semantic mapping which reflects important aspects of the web 2.0 of scientific ideas and perspectives. The knowledge and its inner semantic relation may implicit in textual documents and the minds of a scholar. With the construction of scientific semantic ontology, it is possible that computer can find and filter the required information and knowledge.

Our study focused on developing an ontology-based scientific information (typically for keywords) integration and recommendation system in Web 2.0 circumstance. We try to extract the meaningful information and recommend to web scholars through the constructed system. The rest of the paper is organized as follows. Section 2 describes the related work of keywords recommendation methods and ontology based information management. Section 3 profiles the system design and how they are constructed. Section 4 provides the experimental results and performance evaluation. Section 5 concludes this work and discusses the future studies.

II. RELATED WORK

A. Web 2.0 and Scientific Knowledge Management

Since its first usage of the Web as a mere one-way communication medium, where there was a clear distinction between content providers and content consumers, things have radically changed. Currently, we see a clear trend of users acting not only as content consumers, but also as content providers. Social networks, wikis, blogs, and folksonomies, equipped with newer and more powerful technologies like AJAX, Flash, or RSS, represent what is commonly called Web 2.0 [6]. We will attempt to clarify the concepts behind these terms, and give the reader a view of what they are all about.

The term "Web 2.0" was first used to describe web sites that use technology beyond the static webpages by Darcy DiNucci in 1999, but began its rise in popularity when O'Reilly Media and MediaLive hosted the first Web 2.0 conference in 2004. Providing with more user-interface, software and storage facilities all through the browser, Web 2.0 websites allow users to do more than just retrieve information. This has been called "network as platform" computing. It is characterized with social networking sites, user created web sites, self-publishing platforms, tagging, and social bookmarking. Users can provide the data that is on a Web 2.0 site and exercise some control over that data [7]. Further characteristics of Web 2.0 are: rich user experience, user participation, dynamic content, metadata, web standards and scalability, openness, freedom and collective intelligence [8]. Rich Internet application (RIA), Web-oriented architecture (WOA) and Social Web of Web 2.0 bring with the increasing use of blogs, wikis, and social networking technologies.

Web 2.0 has changed the way of doing research. It led Science 2.0 in academia, which refers to Web 2.0 technologies as the source of the new version in scientific research area. Science 2.0 increases communication and collaboration between scholars, through using computer networking and the Internet. It also suggests the benefits of openness and sharing, regarding scientific knowledge, like papers, research ideas and partial solutions. Scholars use wikis, blogs, video journals, and other collaborative web technologies to share findings, which may include raw data and "nascent theories" online [9].

B. Ontology Based Information Recommendation Methods

Ontology is defined as the "formal, explicit specification of a shared conceptualization" [10]. It is introduced in artificial intelligence field to define the content of domain knowledge, express knowledge, and for communication and sharing. It provides semantic modeling of domain related concepts, attribute, relations and instances for common knowledge sharing and reuse. In context-aware recommender systems, the role of ontology is to help modeling the domain for the tailoring of the right information and services to users and facilitating user-system interaction and system communication with other agents [11]. Cantador, Bellogin, and Castells [12] presented New@hand, a news recommender system which uses semantic technologies to provide personalized and context-aware on-line news recommendation services. Weng and Chang [13] proposed to use ontology and the spreading activation model for research paper recommendation, elevating the system performance and also improving the shortcomings of today's recommendation systems. Their study utilized ontology to construct user profiles and then used to reason about the interests of users. Subramaniyaswamy and Chenthur Pandian [14] proposed a novel approach based on topic ontology for tag recommendation. The proposed approach intelligently generates tag suggestions to blogs. Topic ontology is constructed through enriching the set of categories in existing small ontology called as Open Directory Project in this approach.

Several researchers have explored applications of scientific information recommendation. In order to facilitate more advanced search functions, Tho, Hui and Fong [15] developed a n improved system that employs two techniques: Context-based Cluster Analysis (CCA) and Context-based Ontology Generation frAmework (COGA). Yang [16] conducted research of focusing on developing an ontology-supported information integration and recommendation system for scholars. Xu, Guo, Hao et al. [17] proposed a network based researcher recommendation approach which combines social network analysis and semantic concept analysis in a unified framework to improve the effectiveness of personalized researcher recommendation.

In our study, an ontology based scientific keywords integration and recommendation system was proposed. Through the constructed semantic relation, including

synonymy set and homonymy set, scientific document can be classified more clearly. We try to extract the meaningful information and recommend to web scholars through the constructed system.

III. SYSTEM DESIGN

A. Framework and Services of Keywords Recommendation System

The design framework of the ontology based scientific keywords recommendation system is illustrated in Fig.1. The components of the system are Integration Interface, Service Module, Text Processor, Recommendation Module and Ontology Database. First, the services for scholars include query related keywords recommendation and document keywords recommendation. The former would provide a filtered semantic related keywords list for the search terms scholars entered. This service helps researchers find more professional expression. The other recommendation is for submitted document (e.g. blog) on Web 2.0 sites. The preprocessing of the documents online is supported by Text Processor, which extracts term information from the submitted text. Further, the recommendation module process would calculate term frequency in text, get the semantic related term set from ontology database, add weight on terms, rank the keywords list, and filter the terms to specific requirement.

Ontology database provides complete semantic models for all related keywords. For example, term "ant" may appear in "ant colony optimization" and "ant communication way" having different meanings in two research fields. The ant colony optimization is a heuristic algorithm for combinatorial optimization problems, while ant communication way belongs to biological study. Through separated irrelevant terms from destination returns through field barriers. The keywords ontology extracts topical terms from the online representative academic publications. The semantic relevance of the keywords is built based on WordNet, which is a lexical database grouping English words by sets of synonyms (called synsets) and providing the various semantic relations between these synonym sets.
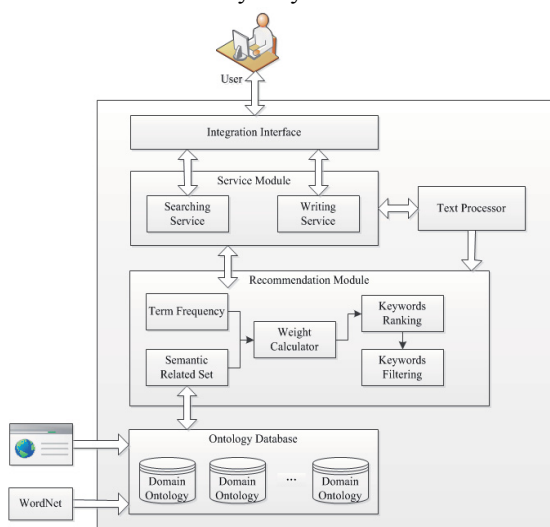
The scientific keywords recommendation system is developed based on semantic ontology. Not only can it provide suggestions of academic search terms, but also it can recommend keywords for Science 2.0 website blogs and the research manuscripts.

B. Scientific Information Integration with Ontology Construction

The role of scientific ontology is to reflect scholars' consensus on a useful way to conceptualize a particular research domain and to organize academic information in semantic mode. Ontology integrates domain knowledge in terms of concepts, properties, relations and axioms. Considering of domain characteristic, scientific objects can be categorized into Researcher, Research Activity, Research Achievement, Research Theme, and Research Organization. In actual scientific area, keywords are the expression of research theme or topic that differentiates studies from others effectively. The scientific ontology construction process in our study is restricted to keywords ontology, shown in Fig. 2. The keywords collection is first extracted from academic articles and other form of publications by domain. Then the process of redundancy reduction is conducted to cut the reduplicate keywords. For example, the difference of word capitalization, space and "-" should be ignored.

After the preprocessing, the semantic similarity of keywords is calculated based on the lexical databases (e.g. WordNet, Wiki, and so on). The semantic similarity of two phrases is computed by (1).

$$S(T_1,T_2) = \delta S_s + (1-\delta)S_r = \delta \frac{s_1 \cdot s_2}{\|s_1\| \cdot \|s_2\|} + (1-\delta)\frac{\|r_1 - r_2\|}{\|r_1 + r_2\|} \quad (1)$$

Where $\delta \leq 1$ decides the relative contributions of S_s, semantic information, and S_r, word order information, to overall similarity computation. According to Li et al. [18], δ is suggested to 0.85. s_i and r_i represent semantic vector and word order vector respectively .

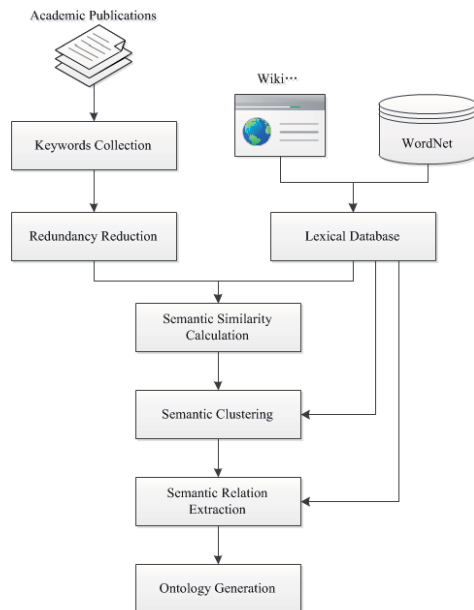Figure 1.   Ontology based keywords recommendation system design

Figure 2.   The process of keywords ontology construction

According to the similarity and semantic dictionary, all keywords would be clustered into synonym sets. Further, the hierarchy with hypernym and hyponym relation, and other correlated relations need to be added in the structure. Finally, the ontology could be generated and visualized in the protégé tool, an open source ontology editor and a knowledge acquisition system. Fig. 3 shows Part of the information retrieval theme taxonomy in scientific topic ontology in protégé. Then, the semantic distances between keywords are all available in ontology.

C. Ontology Based Keywords Recommendation Approach

Recommendation module performs the following tasks: (1) Extract the keywords cluster and related terms set from ontology according to user terms. (2) Calculate the distances of extracted terms and user terms, then add weight on each candidate terms. (3) Ranking the candidates and select top N or α-cut recommendations. The recommendation results are in two forms: alternative keywords and related keywords for searching index and supplementing document information. Fig. 4 shows the process.
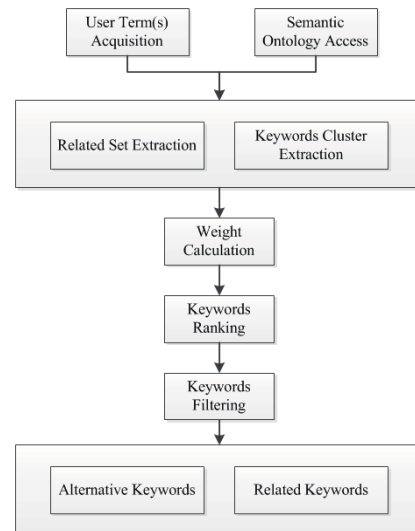
IV. PERFORMANCE VALIDATION

A. Searching Terms Recommendation

Regarding large part of search engines (like Baidu, Google), the related terms are recommend for users. Our experiment on searching terms recommendation is comparing the proposed system and the Baidu, Google search engines. We searched with 50 keywords about web mining obtained from representative articles to examine the recommended terms. The recommended keywords are filtered to be 3 alternative terms and 5 related terms respectively. A sample of the comparison is presented in Table I.

As can be seen from the result, our proposed system can provide semantic correlated keywords efficiently, which is outperforming other search engines. Both Baidu and Google provide some meaningless phrases, which should be ignored. And the distinction between similar phrases and related phrases is quite helpful for searchers.
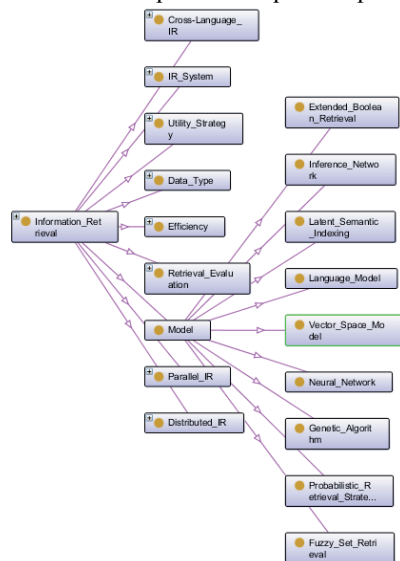
Figure 3.   Part of the information retrieval theme taxonomy

Figure 4.   Ontology based keywords recommendation process

TABLE I.
A SAMPLE OF COMPARING SEARCHING KEYWORDS RECOMMENDATION

User search	Applied system	Recommendation result
web usage	Baidu	usage; usage error 10100; usage data upload; attributeusage; usage error 662; codonusagedatabase; usage error 10027; data usage monitor; codon usage; usage error 10010
	Google	web usage mining; usage; web usage mining discovery and applications of use; web usage mining for web site evaluation; web usage mining survey
	Ontology based system	web usage mining; web usage; data usage (alternative terms)
		general access patterns; customized access patterns; usage profiles; general access logs; web mining (related terms)
social collaboration	Baidu	collaboration; collaboration guy; social; social network; socialbeta; social media; socialpath; social cache; social club; social cn cache
	Google	collaboration; e-collaboration
	Ontology based system	social cooperation; network collaboration; social networks
		online communication; social trust; online forums; share; collaboration software
recommendat-ion system	Baidu	recommendation; system; reboot system now; .utsystemconfig; system pause; system32; system.img; system cls; system app
	Google	recommendation; recommendation system survey; recommendation engine; recommender system; recommendation system java
	Ontology based system	recommendation; recommendation engine; recommender system;
		content based recommendation; collaborative filtering; user profile; e-recommendation; product recommendation

B. Document Keywords Recommendation

The selected data includes 100 papers with 10 topics, i.e., 10 papers for each topic, about knowledge discovery obtained from Google Scholar digital database (http://scholar.google.com.hk/schhp?hl=zh-CN). Firstly, the keywords in papers are removed from the document. Then each paper is submitted to the system for implementing the keywords recommendation. The number of the recommended keywords is set as 5 per article. The comparison of returned results and the removed keywords is conducted with metrics of Precision and Recall, calculated by (2) and (3).

$$Precision = \frac{\left| K_O \bigcap K_R \right|}{\left| K_R \right|} \qquad (2)$$

$$Recall = \frac{\left| K_O \bigcap K_R \right|}{\left| K_O \right|} \qquad (3)$$

where K_O and K_R are the keywords set chosen by author and recommendation result respectively.

The evaluation result of document keywords recommendation is represented in Table II. As is shown, the system could gain about half part of the keywords. More important is that it can provide formalized conceptual expression with ontology available.

V. CONCLUSIONS AND FUTURE WORK

This paper has developed an ontology based scientific keywords recommendation system in Web 2.0 circumstance. The main goal achieved is to extract the meaningful information and recommend to web scholars through the proposed system. The scientific keywords recommendation system is developed based on semantic ontology. Not only can it provide suggestions of academic search terms, but also it can recommend keywords for Science 2.0 website blogs and the research manuscripts. Through the domain ontology, high specific level topics and their corresponding semantic relations are built. The role of scientific ontology is to reflect scholars' consensus on a useful way to conceptualize a particular research domain and to organize academic information in semantic mode.

TABLE II.
PERFORMANCE OF DOCUMENT KEYWORDS RECOMMENDATION

Topic	Precision	Recall
1	0.42	0.49
2	0.56	0.51
3	0.48	0.47
4	0.64	0.55
5	0.66	0.52
6	0.46	0.39
7	0.54	0.43
8	0.52	0.46
9	0.56	0.53
10	0.48	0.47
Average	0.53	0.48

The components of the system are Integration Interface, Service Module, Text Processor, Recommendation Module and Ontology Database. Experimental results and performance evaluation shows that the proposed system provides the effective way to recommend semantic related keywords for scholars.

Currently, the ontology is built by the defined process. The evolution may involve all Web 2.0 scholars, by opening to inner researchers of identical domain. Moreover, other forms of scientific information should be added into this ontology and recommended to scholars to facilitate their online studies.

REFERENCES

[1] Brossard, Dominique and Dietram A. Scheufele. "Science, New Media, and the Public." Science 339, vol. 6115, pp. 40-41, 2013.

[2] Slotta, JamesD and Hedieh Najafi. "Supporting Collaborative Knowledge Construction with Web 2.0 Technologies." In Emerging Technologies for the Classroom, edited by Chrystalla Mouza and Nancy Lavigne, pp. 93-112: Springer New York, 2013. http://dx.doi.org/10.1007/978-1-4614-4696-5

[3] Shum, Simon Buckingham, Enrico Motta and John Domingue. "Scholonto: An Ontology-Based Digital Library Server for Research Documents and Discourse." International Journal on Digital Libraries 3, vol. 3, pp. 237-248, 2000.

[4] Brewster, C., O'Hara, K. "Knowledge Representation with Ontologies: Present Challenge - Future Possibilities." International Journal of Human-Computer Studies 65, vol. 7, pp. 563-568, 2007.

[5] Gruber, T.R. . "The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases." In Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference, edited by R. Fikes J. A. Allen, E. Sandewall 601-602. Cambridge, MA, 1991.

[6] Casteleyn, Sven, Florian Daniel, Peter Dolog and Maristella Matera. "Semantic Web and Web 2.0." In Engineering Web Applications, 293-316: Springer Berlin Heidelberg, 2009.

[7] O'reilly, Tim. "What Is Web 2.0." 2005.

[8] Best, David. "Web 2.0: Next Big Thing or Next Big Internet Bubble." Technische Universiteit Eindhoven, 2006.

[9] Liu, Xiaoyan, Zhiling Guo, Zhenjiang Lin and Jian Ma. "A Local Social Network Approach for Research Management." Decision Support Systems, in press.

[10] Gruber, T. R. "A Translation Approach to Portable Ontology Specifications." Knowledge Acquisition 5, vol. 2, pp. 199-220, 1993.

[11] Buriano, Luca, Marco Marchetti, Francesca Carmagnola, Federica Cena, Cristina Gena and Ilaria Torre. "The Role of Ontologies in Context-Aware Recommender Systems." In Mobile Data Management, 2006. MDM 2006. 7th International Conference on, 80-80: IEEE, 2006.

[12] Cantador, Iván, Alejandro Bellogín and Pablo Castells. "Ontology-Based Personalised and Context-Aware Recommendations of News Items." In Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on, 1, 562-565: IEEE, 2008.

[13] Weng, S.-Sh.., Chang, H.-L. "Using Ontology Network Analysis for Research Document Recommendation." Expert Systems with Applications 34, vol. 3, pp. 1857-1869, 2008.

[14] Subramaniyaswamy, V., Chenthur Pandian, S. "Effective Tag Recommendation System Based on Topic Ontology Using Wikipedia and Wordnet." International Journal of Intelligent Systems 27, vol. 12, pp. 1034-1048, 2012.

[15] Tho, Q.T., Hui, S. C., Fong, A. C. M. "A Citation-Based Document Retrieval System for Finding Research Expertise." Information Processing & Management 43, vol. 1, pp. 248-264, 2007.

[16] Yang, Sheng-Yuan. "Developing an Ontology-Supported Information Integration and Recommendation System for Scholars." Expert Systems with Applications 37, vol. 10, pp. 7065-7079, 2010.

[17] Xu, Yunhong, Xitong Guo, Jinxing Hao, Jian Ma, Raymond Y. K. Lau and Wei Xu. "Combining Social Network and Semantic Concept Analysis for Personalized Academic Researcher Recommendation." Decision Support Systems 54, vol. 1, pp. 564-573, 2012.

[18] Li, Y.H., McLean, D., Bandar, Z.A., O'Shea, J.D., Crockett, K. "Sentence Similarity Based on Semantic Nets and Corpus Statistics." IEEE Transactions on Knowledge and Data Engineering 18, vol. 8, pp. 1138-1150, 2006.

AUTHORS

Na Xue is currently a Ph. D candidate in management science and engineering at School of Economics and Management from Beihang University, Beijing, CO 100191 China (e-mail: xuena12530@163.com).

Su ling Jia is a Professor at School of Economics and Management, Beihang University, Beijing, CO 100191 China (e-mail: jiasuling@126.com).

Jin xing Hao is an Assistant Professor at School of Economics and Management, Beihang University, CO 100191 China (e-mail: gordonhao@gmail.com)..

Qiang Wang is an Associate Professor at School of Economics and Management, Beihang University, Beijing, CO 100191 China (e-mail: wang6965@sina.com).