Finding Hidden Semantics Behind Reference Linkages: An Ontological Approach for Scientific Digital Libraries*

Peixiang Zhao[1], Ming Zhang[2], Dongqing Yang[2], and Shiwei Tang[2]

[1] Department of Systems Engineering and Engineering Management,
The Chinese University of Hong Kong,
Hong Kong SAR, China
pxzhao@se.cuhk.edu.hk

[2] Department of Computer Science and Technology, Peking University,
Beijing, China
{mzhang, ydq}@db.pku.edu.cn
tsw@pku.edu.cn

Abstract. The contents and topologies of inter-document linkages, such as citations and references among scientific literature, have received increasing research interests in recent years. Some technologies have been fully studied and utilized upon this meaningful information to improve the organization, analysis and evaluation of scientific digital libraries. In this paper, we present a CiteSeer-like system to access scientific papers in computer science discipline by reference linking technique. Moreover, implicit semantics behind reference indices are mined and organized to improve accessibility of scientific papers. In order to model scientific literature and their interlinked relationships, we develop a domain-specific ontology to analyze contents and citation anchor context of scientific papers. Compared with abstract of a specific paper written by authors themselves, we introduce an automatic summary generation algorithm to create objective descriptions from other scholars' perspectives based on the ontology. Semantic queries can also be asked to discover interesting patterns in scientific libraries in order to provide a comprehensive and meaningful guidance for users.

1 Introduction

With the rapid development of Internet and the increasing ripeness of the Web, more and more scientific papers appear on the Web in digital form instead of paper-based form. These digitized scientific documents have greatly facilitated the Web to be an efficient repository of up-to-date information. However, as the availability of scientific literature greatly improves, the inability of people to disseminate, share and profitably utilize such a large amount of information becomes more and more severe. Published scientific papers available on the Web are widely spread and are often poorly organized, neither comprehensively indexed nor interlinked in terms of logical correlations among them. People get limited supports in searching, reviewing, and

* This research is funded in part by NSFC grant 90412010 as well as grant 60221120144.

L. Zhou, B.C. Ooi, and X. Meng (Eds.): DASFAA 2005, LNCS 3453, pp. 699–710, 2005.
© Springer-Verlag Berlin Heidelberg 2005

analyzing scientific literature from academic perspectives. This problem is now becoming an important research issue both in computer science and in digital libraries.

One original approach for managing such a huge volume of scientific literature is by reference linking different papers from bibliographic perspectives. Reference indices are useful for a number of purposes, such as literature dissemination, search, analysis and evaluation. Fig.1 shows main components of a reference index and a sample of reference linking. The left diagram gives a link topology among different scientific papers denoted by node *A*, *B*, *C*, *D*, *E*, *F* and *G*. The right diagram shows article *D* cites article *G*. Components of a reference index between *D* and *G* are citation anchor context in *D* and the corresponding bibliographic entry in the reference section of *D*. Through a reference link from "citing" papers to the "cited" one, users can find citation patterns and relationships among different documents. By navigating backward and forward though reference indices, users can promptly find a series of papers and perceive a thorough understanding of related research. The context of a citation in citing papers is quite illuminating in judging the motivation of reference, the contributions of the cited paper and the usefulness of a paper for a given query [11]. Reference indices are now widely used in reference linking [4], link analysis [3,7], hypertext and web mining [6], text classification [12] etc. The potential usefulness of reference indices contained in scientific literature has now been widely convinced and related services are provided in a range of applications, especially in scientific digital libraries.

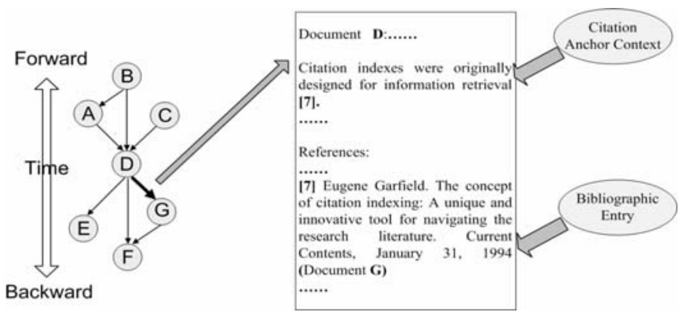


Fig. 1. A reference linking sample: from the time perspective

In resent years, several digital libraries have established large repositories of scientific literature, such as ISI SCI®, CiteSeer.IST [10], CORA [2], and ML Papers [13] etc. These projects pay much attention to interlinking different scientific papers via reference indices. However, few of them richly utilize semantic information hidden behind reference linkages. HITS [9,3] and PageRank [17] algorithms are widely applied to identify "important" web pages or papers in the web and digital libraries by analyzing link topologies of hyperlink or citation graph, but both of them neglect to analyze contents of web pages and scientific papers. In this paper we present a reference linking prototype for automatically linking scientific literature in computer science discipline. We develop a domain-specific ontology that models scientific literature and interlinked reference indices in order to identify semantics behind reference linkages. As to a specific scientific paper, we use descriptions derived from citation anchor context of citing ones to automatically generate summary which includes information of research theme and motivation, research topics, research background, research impact and applied fields etc. Compared with the abstract written by authors themselves, the summary is objective from different scholars' perspectives. Thus it is quite helpful for researchers to understand the literature better. Further reasoning upon the ontology is also provided by semantic query form in order to reveal new facts in scientific digital libraries. In this way, users can perceive a comprehensive understanding of a specific research domain, but not merely one scientific paper.

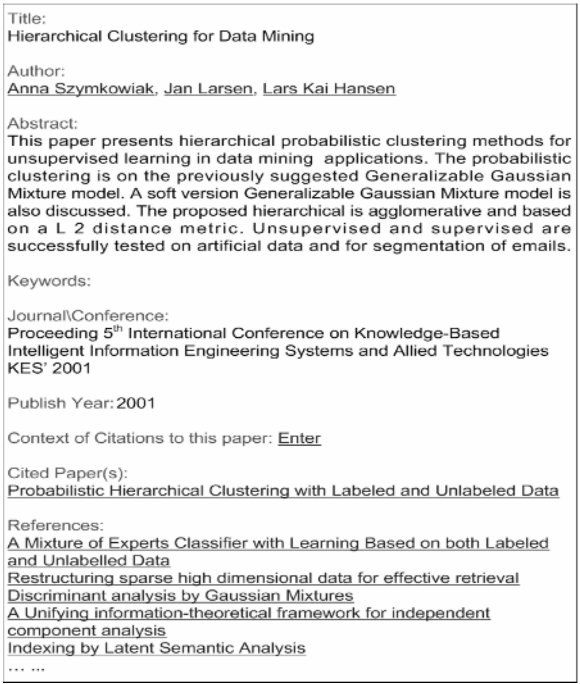
The reminder of this paper is organized as follows. Section 2 describes how to interlink scientific literature with metadata information. In section 3 there is a detailed description of the domain-specific ontology construction. Based on this ontological knowledge base, section 4 comments on how to automatically generalize summary information of a scientific paper and to provide further reasoning upon the ontological knowledge base by semantic queries. Section 5 shows experiments and evaluations of the prototype system. The final section presents conclusions of our work.

2 Reference Linking with Metadata

Reference linking means turning references within a scientific paper into "live reference" so that you can follow them in citing article to other accessible cited ones [4]. In order to facilitate reference linking in scientific literature, the first task is to extract metadata precisely and automatically from papers. Our metadata definition of scientific literature is derived form Dublin Core [5], which consists of title, author information, abstract, keywords, content of paper, bibliographic information, citation anchor context and appendix etc. Due to the inconsistency of metadata formats in scientific literature, we use information extraction techniques to improve the parsing of documents.

We use heuristic rules and regular expression matching technique to extract metadata from scientific papers. Extraction rule database is applied to accommodate various bibliographic styles appeared in scientific literature. We also use databases of author name, journal/conference name and domain name to help identify metadata. Font information and layout clues are quite helpful to determine specific metadata, such as title of a scientific paper. Different granularity strategies during metadata extraction are applied. For metadata such as title, abstract, content etc., coarse granularity strategy is used because these metadata could be extracted within one path. Metadata such as author information and bibliographic information should be extracted by fine granularity strategy in order to identify subfields of metadata. As shown in Fig.1, a reference index has two components: the detailed reference information in the bibliographic entry and citation anchor context (the sentences occurring near the citation tag) in the content of citing papers. Thus a reference index should be matched with its corresponding citation anchor context for further analysis.

After metadata extraction and reference/context matching, metadata of scientific papers are stored in citation database. Fig.2 shows the detailed metadata information of a scientific paper "*Hierarchical Clustering for Data Mining*" retrieved by a standard *SQL* query. The "Cited Paper(s)" section lists a paper "*Probabilistic Hierarchical Clustering with Labeled and Unlabeled Data*" who cites "*Hierarchical Clustering for Data mining*" and the "Context of citations to this paper" section lists corresponding citation anchor context. The "References" section lists scientific papers cited by "*Hierarchical Clustering for Data mining*"


Title:
Hierarchical Clustering for Data Mining

Author:
Anna Szymkowiak, Jan Larsen, Lars Kai Hansen

Abstract:
This paper presents hierarchical probabilistic clustering methods for unsupervised learning in data mining applications. The probabilistic clustering is on the previously suggested Generalizable Gaussian Mixture model. A soft version Generalizable Gaussian Mixture model is also discussed. The proposed hierarchical is agglomerative and based on a L 2 distance metric. Unsupervised and supervised are successfully tested on artificial data and for segmentation of emails.

Keywords:

Journal\Conference:
Proceeding 5th International Conference on Knowledge-Based Intelligent Information Engineering Systems and Allied Technologies KES' 2001

Publish Year: 2001

Context of Citations to this paper: Enter

Cited Paper(s):
Probabilistic Hierarchical Clustering with Labeled and Unlabeled Data

References:
A Mixture of Experts Classifier with Learning Based on both Labeled and Unlabelled Data
Restructuring sparse high dimensional data for effective retrieval
Discriminant analysis by Gaussian Mixtures
A Unifying information-theoretical framework for independent component analysis
Indexing by Latent Semantic Analysis
… …

Fig. 2. Metadata of a scientific paper "Hierarchical Clustering for Data mining"

3 The Domain-Specific Ontology Construction

Ontology is the study of "things that exist" that began as a branch of philosophy and is now popular in the field of knowledge management [8]. We are developing an ontology that models scientific literature and their interlinked relationships toward producing an indexing and evaluation system for scientific digital libraries, as shown in Fig.3. By analyzing contents of cited paper and citation anchor context of citing ones, we automatically identify concepts with the aid of the ontological knowledge base. We also extract implicit claims concerning cited paper's motivations, contributions and relationships to corresponding research issues by analyzing citation anchor context of citing papers,. The interlinked concepts, together with semantic claims are modeled to form ontology for scientific literature in computer science discipline. This ontological knowledge base is powerful to intelligently communicate, analyze and reason over concepts and knowledge of scientific literature.

3.1 Topic Distillation by WordNet and Ontological Knowledge Base

We want to build ontology that models concepts in scientific literature and their relationships described in citation anchor context. Thus we have to extract concepts, namely topics from scientific papers. We analyze title and keywords of cited paper, together with descriptions in citation anchor context of all citing papers to distill research topics. Because title and keywords are written by authors themselves whereas citation descriptions by other scholars and researchers, a combination of analyses both from subjective perspective and from objective one is proved to be convincing.

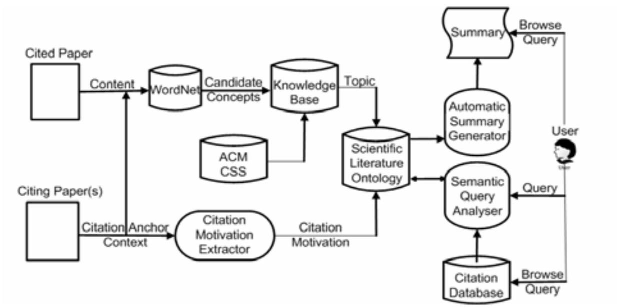


Fig. 3. The construction process of the ontological knowledge base that models scientific literature in computer science discipline

As to keywords of cited paper, we simply separate each keyword from the other. Since title and citation anchor text are always sentences describing cited paper, and the most common type of concept is a sequence of proper nouns or noun phrases, we use WordNet [14], a successful concept ontology, to parse contents of these sentences. After proper stemming and sense-tagging by WordNet, nouns and noun phrases are extracted from sentences. These nouns and noun phrases, together with the separated keywords, are candidate concepts of cited paper for further analysis.

While WordNet is a general purpose concept ontology, it can not be expected to provide exhaustive coverage of concepts in some specific domains. We develop an ontological knowledge base in computer science discipline to help distill research topics from candidate concepts. This knowledge base is a concept hierarchical tree-like ontology derived from ACM CCS (Computing Classification System) [1], an existing knowledge base in computer science discipline. Each candidate concept of paper will be mapped on the node of the knowledge base and those with a complete match are topics of related scientific papers. We calculate weights for each topic that represents its importance as a descriptor for paper. Concept weight is calculated using the metric based on TFIDF, a standard term-weighting measure from the information retrieval research community.

$$w_t = \frac{n_t}{1 + \log N_t} \tag{1}$$

Where w_t is the weight of the concept *t* completely matched with corresponding node in the knowledge base. n_t is the number of times concept *t* appeared in title, keywords of cited paper and in citation anchor context of citing papers. N_t is the number of documents concept *t* appears.

3.2 Ontological Knowledge Base Construction

As mentioned in section 3.1, we develop an ontological knowledge base in computer science discipline to help extract research topics from scientific literature. Based on ACM CSS, our knowledge base has a hierarchical tree-like structure. There are two relations between concepts in knowledge base: the "*Is-a*" relation in the same categorization sub-tree and the "*Similar*" relation between different categorization sub-tree. As shown in Fig.4, the concept "*Statistical database*" has a "Is-a" relation with "*Database Management*" and has a "Similar" relation with "*Probability & Statistics*" in another categorization sub-tree. A hash table is built to facilitate concept matching process between candidate concepts and nodes in knowledge base.

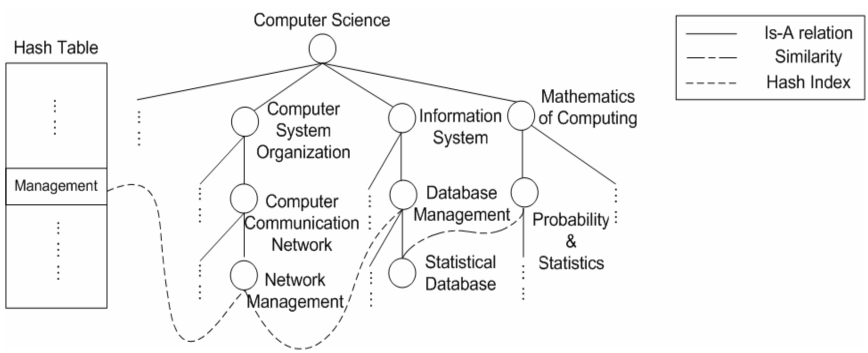


Fig. 4. The ontological knowledge base in computer science discipline

Upon the knowledge base, we define operations to identify semantic relations between different nodes. From a concept node in knowledge base, we could find more general or specific concepts by ***Parent*** operation or ***Child*** operation respectively and those similar concepts could also be retrieved by ***Similar*** operation. ***Sim*** operation provides anther approach to measure similarities of different concepts by calculating semantic distances in the knowledge base. In the following equations, X, Y and Z denote concepts of the knowledge base.

$$\textbf{\textit{Parent}}\,(X) = \{\,Y \mid \text{where } X \text{ has a direct "Is-A" relation with } Y\} \tag{2}$$

$$\textbf{\textit{Child}}\,(X) = \{\,Y \mid \text{where } Y \text{ has a direct "Is-A" relation with } X\} \tag{3}$$

$$\textbf{\textit{Similar}}\,(X) = \{Y \mid \text{where } Y \text{ has a direct "Similar" relation with } X\} \tag{4}$$

Dist *(X, Y)* **=** number of "Is-a" relation links between *X* and *Y* in the same categorization sub-tree (5)

Sim *(X , Y)* = ***Dist*** *(X, Z)* + ***Dist*** *(Y, Z)*, where Z is the nearest parent node of both X and Y (6)

Based on the ontological knowledge base mentioned above, we could conveniently extract research topics from scientific papers in computer science discipline. In the next section, we will analyze citation anchor context to find citation motivations and relationships between cited paper and citing ones.

3.3 Citation Motivation Extraction and Analysis

Citation anchor context of citing papers are "meaningful" text which can provide detailed descriptions and evaluations to the cited one. It is quite helpful to find out "why" and "how" a particular scientific paper is cited and relationships to other papers in the literature. We extract citation motivations from citation anchor context and classify them into several types in order to assign semantic meanings to interlinked reference indices in scientific literature. [18] proposed 15 categories for the reasons of citation, but we classify reference motivations into the following 5 categories:

- Providing background knowledge
- Identifying methodology
- Pointing out problems or drawbacks
- Indicating or predicting future research
- Others

Similar to [15], we make rules for each category and extract reference motivations based on cue phrases. When a cue phrase is identified in citation anchor context, we classify the citation motivation of this reference index into the category which the cur phrase belongs to. For example, if the phrases in citation anchor context such as "make use of", "present", "applied to" etc. are identified, corresponding reference indices are classified to the "Identifying methodology" category; if the phrases with "little influence", "inconsistent with", "raise problems" etc. are identified, the reference indices are classified to the "Pointing out problems or drawbacks" category. For each citation motivation, we assign an attribute to identify attitudes of authors toward the cited paper: 1 for positive attitude, -1 for negative attitude and 0 if the attitude can not be identified. By analyzing content of citation anchor context, especially those verbs mentioned above, and adjectives or adverbs, such as "extensively", "efficiently", "difficultly" etc., attributes of citation attitude could be extracted which implies inclinations of authors whether to recommend or criticize related research described in the cited paper.

After citation motivation extraction and analysis, semantic meanings are identified and assigned to reference indices in scientific literature. It offers intellectual linkage among different papers and it is helpful for users to learn more from citing papers about the cited one by descriptions and evaluations from other researchers' perspectives. It also provides us with meaningful linkage among concepts of our own ontological knowledge base. Further analysis will be applied upon the ontological knowledge base to infer interesting citation/reference patterns in scientific literature.

4 Analyses and Reasoning Upon the Ontology

4.1 Automatic Summary Generation

As shown in Fig. 3, based on research topics extracted from scientific papers and citation motivations from interlinked reference indices, we develop an ontological knowledge base to model scientific literature and their relationships. This knowledge base is efficient to communicate, analyze and reason over concepts and knowledge of scientific literature. One of the most important applications upon it is automatic summary generation.

A summary of a scientific paper is very helpful and instructive for readers to know what has been studied. As writing a summary or a survey by manual work is quite time-consuming, it is desirable to generate comprehensive summaries for scientific papers automatically. In our prototype system, summary of a scientific paper is composed of four components:

- research theme and motivation
- research topics
- research background
- research impact and applied fields

Based on the ontology of scientific literature in computer science discipline, we apply automatic summary generation algorithm to get each component of summary for scientific papers. As to the component "*research theme and motivation*", we have got results by analyzing citation anchor context of citing papers in section 3.3. The results are grouped by different citation categories and ordered by attributes of authors' attitudes. Users could follow hyperlinks to examine detailed descriptions in citation anchor context of each citing papers. As to the component "*research topics*", we have got results by topic distillation with the aid of WordNet and ontological knowledge base in section 3.2. Research topics are ordered by concept weight defined in equation (1). Because research topics are matched with concepts in the knowledge base, we can apply operations defined from equation (2) to (6) upon corresponding concepts of the knowledge base to get additional information for topics of scientific papers. For example, while referring to subordinate topics of "*Information Systems*", we can get research issues such as "*Database Management*", "*Information Retrieval*", "*Digital Library*" etc. While referring to similar topics of "*File*", we can get topics such as "*File Systems Management*", "*Database Management*" etc., which may share research similarities with the topic "*File*" but are quite different issues in some other research background. As to the component "*research background*" and "*research impact and applied fields*", we simply organize research topics of cited papers and citing ones, respectively. The topics are also grouped and ordered by concept weight. Fig. 5 presents a summary automatically generated by our algorithm. The summary consists of four components mentioned above which gives a comprehensive description for the scientific paper "*Melodic matching techniques for large music databases*". Users can follow hyperlinks to get detailed information of related research issues and corresponding scientific papers.

Title:	
Melodic matching techniques for large music databases	(7 citations)
Summary:	
Theme and Motivation:	
Identifying methodology	(3 papers)
Providing background reading	(1 paper)
Others	(2 papers)
Pointing out problems or drawbacks	(1 paper)
Research Topic:	
Music database	(3 papers)
Information Retrieval	(2 paper)
Searching	(2 papers)
Research Background:	
Music database	(3 papers)
Information retrieval	(3 papers)
Pattern Recognition	(3 papers)
Algorithm	(2 papers)
Artificial Intelligence	(1 paper)
Impact and Applied Fields:	
Music Database	(4 papers)
Information retrieval	(2 papers)
Data Structure	(1 paper)

Fig. 5. Summary of the paper "Melodic matching techniques for large music databases"

4.2 Reasoning by Semantic Queries Upon the Knowledge Base

In addition to automatic summary generation, the ontological knowledge base also makes it possible to infer knowledge in scientific literature. In our prototype, the ontological knowledge base enables discoveries of implicit information by semantic queries which are described as OWL QL [16] query patterns, namely a set of triples of the form *(<property> <subject> <object>)*. Each triple is mapped to several operations of the ontological knowledge base or standard SQL statements on the citation database. Our prototype system provides limited triples describing the most common operations upon the ontology and the citation database. For example, to find scientific papers a research organization has published, we can follow:

Query: ("Scientific papers published by research organization A")

Query Pattern: {(is-author ?a ?p)(work-for ?a ?o) (equal ?o "A")}

Must-Bind Variables List: (? p)

May-Bind Variables List: ( )

Don't-Bind Variables List: (? a)

Where variables *a* stands for author names, *p* stands for scientific papers and *o* stands for research organizations. And, withal, *is-author*, *work-for* and *equal* etc. are semantic queries supported by our prototype system.

Taking advantage of the expressive power of the ontological knowledge base built upon scientific literature, the prototype system can provide structural queries, such as queries asking of super-class, subclasses and similar classes of a given concept in computer science. All these structural queries are formulated to operations upon the ontological knowledge base. We also provide complicated queries involving concepts not expressible in the ontological knowledge base such as "*Which organizations are research communities in information retrieval research domain?*" or "*Who are noted authorities on data mining?*" These complicated queries are constructed by those system-supported query triples. Users can write their own semantic queries in OWL-QL query patterns by means of combining different system-supported queries into more complicated and comprehensive ones. Heuristic rules are also provided to solve analytical queries. For example, to answer the query "*which research subfields are quite relevant to digital library research?*" We define that "*quite relevant*" subfields to "*digital library*" are those research topics with a similarity link to the node "*digital library*" in the ontological knowledge base or research topics whose similarity distances to the node "*digital library*" measured by *Sim* operation are less than or equal to 2. The number 2 represents user definable thresholds.

5 Experimental Evaluation

In our prototype system, we locate and download over 10,000 scientific papers in computer science discipline from Internet by web crawler. After preprocessing, metadata extraction and reference/context matching, metadata information is stored in citation database implemented by PostGreSQL. In order to improve the accessibility of reference linking, documents without full text that can not be parsed or those with less than 3 in-degrees are eliminated from database. There are finally 7,973 scientific papers in our citation database.

The effectiveness of reference linking in scientific literature heavily relies on precision of metadata extraction. We do several experiments to evaluate our extraction technique and final extraction results. We choose 1430 scientific papers in CiteSeer.IST. After preprocessing, 293 papers with errors are eliminated from test beds. The metadata extraction results are shown in Table 1.Our extraction precision is a litter higher than CiteSee.IST and Opcit [4].

Table 1. Metadata extraction precision results of sceintific articles

Automatic Metadata Extraction	**Metadata of Scientific Papers**						
	Title	Author Information	Abstract	Keywords	Content	Reference	Appendix
Extraction Precision	92.1%	87.8%	98.9%	100%	100%	83.3%	100%

While extracting topics from scientific literature described in section 3.1, one of the important steps is to distill candidate concepts with the aid of the ontological knowledge base. However, there is a possibility that some candidate concepts are not perfectly matched with those appeared in knowledge base. We do an experiment to investigate how well candidate concepts from titles, keywords and citation anchor context match with concepts in the ontological knowledge base. We choose 200 scientific articles from ML Papers, 150 for training and 50 for testing. We manually choose concepts from title, keywords and reference anchor context in the training paper set in order to train our ontological knowledge base, especially in machine learning research field. Fig.6. shows concept matching results in the test bed. Each left column shows numbers of concepts that should be extracted and corresponding right column shows numbers of concepts that are perfectly matched and extracted by the ontological knowledge base.

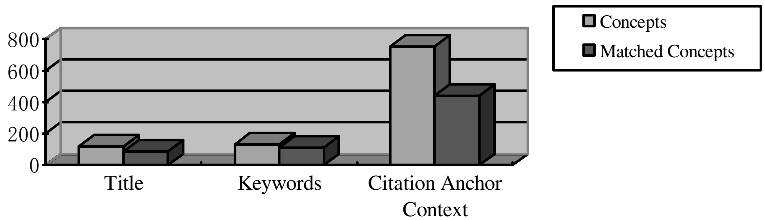


Fig. 6. Concept matching results in the test bed

As shown in Fig.6, concepts matching precisions from title, keywords and citation anchor context sections of scientific papers in the test bed are 74.8%, 82.3% and 58.5%. Compared with concept matching precisions in title and keywords section, that in citation anchor context is low. That is partially because the ontological knowledge base does not contain newly emerging concepts from citation anchor context and some of keywords fall out of machine learning classification sub-trees.

6 Conclusions

In Recent years the Web has developed to be a medium for scientific literature dissemination, retrieval and evaluation. However, experience show that hundreds of thousands of scientific papers are neither comprehensively indexed nor interlinked in terms of citation/reference semantics. In scientific digital libraries domain, reference indices are increasingly being applied to improve the ability of management, analysis and evaluation of scientific literature. In this paper we investigate the usage of a domain-specific ontology to improve semantic linking of scientific literature via reference indices among different scientific papers. The ontological knowledge base is consisted of concepts extracted from papers and reference motivations derived from citation anchor context. Moreover, we utilize automatic summary generation algorithm upon the ontological knowledge base to give an objective description for cited paper from scholars' perspectives. Further reasoning is also provided by system-supported or user-defined semantic queries upon the knowledge base in order to reveal new facts and interesting patterns in scientific libraries.

Our prototype system is implemented and some experiments are conducted. We are now encouraged to apply ontology techniques in scientific digital libraries and further research in this field is ongoing.

Acknowledgements

The authors would like to thank Prof. Jian Pei in Department of Computer Science and Engineering, University of Buffalo, the Sate University of New York for his comments on an early version of this paper.

References

[1] ACM CCS. http://www.acm.org/class/
[2] Andrew Kanchites McCallum, Kamal Nigam, Jason Rennie and Kristie Syemore. "Automating the construction of Internet portals with machine learning". *Information Retrieval Journal*. Volume 3. Pages 127 -163. 2000.
[3] Ding C., Zha H., He X., Husbands P., and Simon H. "Analysis of hubs and authorities on the web". Lawrence Berkeley Nat'l Lab Tech Report 47847(www.nersc.gov/~cding /hits.ps). 2001
[4] Donna Bergmark. "Automatic extraction of reference linking information from online documents". Technical Report TR 2000 -1821, Cornell Computer Science Department. October 2000
[5] Dublin Core Metadata Initiative http://purl.oclc.org/dc/
[6] Flake, G. W., Lawrence, S., and Giles.C.L. "Efficient identification of web communities". *Sixth ACM SIGKDD International conference on Knowledge Discovery and Data Mining*. Pages 150-159, 2000
[7] Gerard Salton. "Automatic indexing using bibliographic citations". *Journal of Documentation*, Volume 27. Pages 98-110. 1971
[8] Guarino, N. "Formal Ontology and Information Systems". in *Proc. FOIS'98 trento*. Italy. 6-8 June
[9] Jon Kleinberg. "Authoritative sources in a hyperlinked environment". In *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*. 1998
[10] Kurt D. Bollacker, Steve Lawrence, C. Lee. Giles. "CiteSeer: an autonomous web agent for automatic retrieval and identification of interesting publications". *Proceedings of 2nd International Conf. on Autonomous Agents. ACM Press*. Pages 116-123. May 1998.
[11] Lempel R. and Moran S. "SALSA: stochastic approach for link-structure analysis and the TKC effect". *ACM Trans. Information Systems*. Volume 19. Pages 131-160. 2001
[12] Lu Q. and Getoor L. "Link-based classification". In *Proc of ICML-03*, 2003
[13] ML Papers. http://www.ai.mit.edu/people/ayn/cgi/vpapers
[14] Miller G.A., Beckwith R., Felbaum C. Gross D. and Miller K. "Introduction to WordNet: An On-line Lexical Database". *International Journal of Lexicography*. Volume 3. Pages 235-244. 1990
[15] Nanba, H., and Okumura M. "Towards Multi-paper Summarization Using Reference Information". *Proceedings of the 16th International Joint Conferences on Artificial Intelligence (IJCAI-99)*. Pages926-931. 1999
[16] R. Fikes, P.Hayes and I. Horrocks. "OWL-QL – a language for deductive query answering on the semantic web". Technical Report KSL-03-14, Knowledge Systems Lab, Stanford University, CA, USA. 2003
[17] Sergey Brin and Lawrence Page. The Anatomy of a Large-scale Hypertextual Web Search Engine. In *the Seventh International World Wide Web Conference*. 1998.
[18] Weinstock, N. "Citation indexes, in Kent A (Ed.)". Encyclopedia of Library and Information Science, New York. Pages 16-41. 1971.