Scholar Metadata and Knowledge Generation With Human And Artificial Intelligence

Xiaozhong Liu and Chun Guo
Department of Information and Library Science, School of Informatics and Computing, Indiana University Bloomington, 1320 East 10th Street, Li 011, Bloomington, IN 47405. E-mail: {liu237, chunguo}@indiana.edu

Lin Zhang
Transportation Management College, Dalian Maritime University, No. 1, Linghai Street, Ganjingzi District, Dalian, 116026, P.R. China. E-mail: joslin@dlmu.edu.cn

Scholar metadata have traditionally centered on descriptive representations, which have been used as a foundation for scholarly publication repositories and academic information retrieval systems. In this article, we propose innovative and economic methods of generating knowledge-based structural metadata (structural keywords) using a combination of natural language processing-based machine-learning techniques and human intelligence. By allowing low-barrier participation through a social media system, scholars (both as authors and users) can participate in the metadata editing and enhancing process and benefit from more accurate and effective information retrieval. Our experimental web system ScholarWiki uses machine learning techniques, which automatically produce increasingly refined metadata by learning from the structural metadata contributed by scholars. The cumulated structural metadata add intelligence and automatically enhance and update recursively the quality of metadata, wiki pages, and the machine-learning model.

Structural Keywords

*<Research Question>*Metadata Generation*</Research Question>*

*<Methodology>*Human Intelligence, Artificial Intelligence, Natural Language Processing (NLP)*</Methodology>*

*<Dataset>*Scholar Publication*</Dataset>*

*<Evaluation>*User Evaluation, Cross-Folder Validation*</Evaluation>*

Introduction

The ability to discover, integrate, and reuse relevant scholarly output from prior studies is critical for innovative research (Shotton, Portwin, Klyne, & Miles, 2009). This ability is largely facilitated by the metadata representing such scholarly output. Metadata have traditionally centered on descriptive representation by title, author, publisher, subject keywords, and other attributes of scholarly output (e.g., the Dublin Core Metadata Element Set). Descriptive metadata, however, are rapidly becoming increasingly inadequate as the complexity and volume of scholarly output grow. Database vendors have developed innovative mechanisms to address these new challenges. Examples include the Association for Computing Machinery (ACM) Digital Library's (http://dl.acm.org/) reference-linking capabilities and biomedical research journals that provide structured abstracts to facilitate search and browsing. These developments, though sporadic and varying in degree of sophistication, have enhanced the retrieval and use of scholarly publications by providing nontraditional descriptive metadata.

The goal of this article is to describe economical methods of creating novel structural (domain-knowledge-based) keywords, a kind of structural metadata, through the integration of human and artificial intelligence. An innovative type of structural metadata, the structural keyword, is a machine-readable device developed to facilitate knowledge retrieval.

Structural metadata (Liu & Qin, in press) can be defined as a framework that scientific papers use to convey an argument or report research results to their "persuasive community" (Allen, Qin, & Lancaster, 1994). For instance, the section headings in a research paper are typical structural elements that can be used as structural metadata. From a rhetoric point of view, scientists must persuade their community through their papers that their research methods are sound and that their results are significant, valid, and reliable. It is agreeable among researchers that a research paper

Received April 24, 2012; revised March 16, 2013; accepted April 26, 2013

 • Published online 22 February 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23013

should clearly describe the research question, methods, and data/materials used to study the problem as well as the results and subsequent discussion. Although these are common structural elements in scientific writing, variations exist between individuals and between disciplines. Such structural elements not only form a persuasive framework for authors to make their argument but also present potentially useful metadata for in-depth representations of research papers. This means that although some structural elements are common across communities, others are specific to individual fields. For example, "research question" is one of the structural concepts used in almost all scientific publications regardless of research domain, but among the structural concepts typically appearing in an information retrieval publication ("research question," "methodology" [sometimes called an "algorithm"], "data set," and "evaluation"), the concept "evaluation" may not be a common structural component in other domains' publications. Although some publications in the information retrieval domain may not strictly follow this rhetorical structure, the fundamental domain structural keywords can assist users in analyzing the microlevel semantics within the paper and identifying relations among different papers. Structural metadata exist not only in the section headings but also the text body of a paper and other descriptive metadata. For this article, we enhanced classical keyword metadata to structural keywords by using human and artificial intelligence.

As an example, one structural keyword of this article noted in the "structural keywords" synopsis (i.e., "<Research Question> Metadata Generation</Research Question>") illustrates the research question, and "<*Dataset*>Scholar Publication</*Dataset*>" describes the data set. Indexing structural keywords is useful for knowledge-based retrieval systems. For example, our prototype knowledge-retrieval engine (Guo, Chinchankar, & Liu, 2012), as Figure 1 shows, can support knowledge-based queries such as "*Methodology: Support Vector Machine; Dataset: TREC QA; Evaluation: NDCG*" for the "information retrieval" domain. Furthermore, as Figure 1 shows, the system can recommend structural keywords for natural language queries and provide knowledge feedback (on the right side) to help users better understand retrieved results or to update their initial query from a knowledge perspective. These innovative features are based on the structural keyword implementation.

However, structural keyword generation for large numbers of publications is a demanding task. Although the cost of employing domain experts is unacceptably high, fully automatic machine learning classifier and natural language processing (NLP) approaches (artificial intelligence) may suffer from low accuracy and training-data sparseness problems. It is now vital to find ways to effectively generate structural keywords for large numbers of publications at a low cost. In this research, we propose a hybrid method of integrating human and artificial intelligence. By allowing low-barrier participation through an experimental social media system, ScholarWiki users (both as authors and users) can participate in the knowledge and metadata editing and enhancing process, which automatically "triggers" the evolution of the machine learning model used to enhance metadata quality. Unlike other social media systems, ScholarWiki allows users to make nonlinear contributions to the metadata repository, which means that creating/editng one publication's metadata can trigger the comprehensive improvement of all other publications' metadata. By using the information retrieval domain as a case study, our experiment shows that artificial intelligence plus crowd-sourced user feedback via the ScholarWiki system is effective in creating high-quality structural keyword metadata.

In the remainder of this article, we (a) review relevant literature and methodologies, (b) introduce the artificial and human intelligence–based structural keyword creation methodology, (c) describe the experiment in the information retrieval domain using 20,799 publications and one graduate-level information retrieval class, (d) evaluate our work by comparing different approaches and features, and (e) discuss the contributions and limitations of our work.

Previous Research

Metadata representation facilitates finding, identifying, selecting, and obtaining information objects (International Federation of Library Associations, 1998). In its short history, metadata research has split into two camps with differing perspectives and paradigms: the description paradigm in library and information science and the processability and executability paradigm rooted in computer science (Zeng & Qin, 2008). Research on metadata representation and generation during the past few decades has drawn techniques and methods from a wide variety of research fields, including NLP, machine learning, classification, and ontology. This study focuses on metadata representation and generation. In this section, we review these two lines of research.

Structural Metadata Representation

The proliferation of cyberinfrastructure-enabled research environments calls for more powerful and effective metadata representation methods to address information discovery challenges. The rapid growth of scholarly publications has highlighted the need for structural metadata, an emerging approach to structuring publication information. The basic assumption is that certain types of domain-specific publications adhere to a predictable knowledge structure. Understanding this structure may promote further advances for managing NLP or information retrieval applications (Y. Guo et al., 2010; Lin, Karakos, Demner-Fushman, & Khudanpur, 2006). More recently, Evans and Foster (2011) noted the importance of metaknowledge by integrating structural information across different publications in terms of explicit knowledge, implicit knowledge (i.e., implicit

DOI: 10.1002/asi

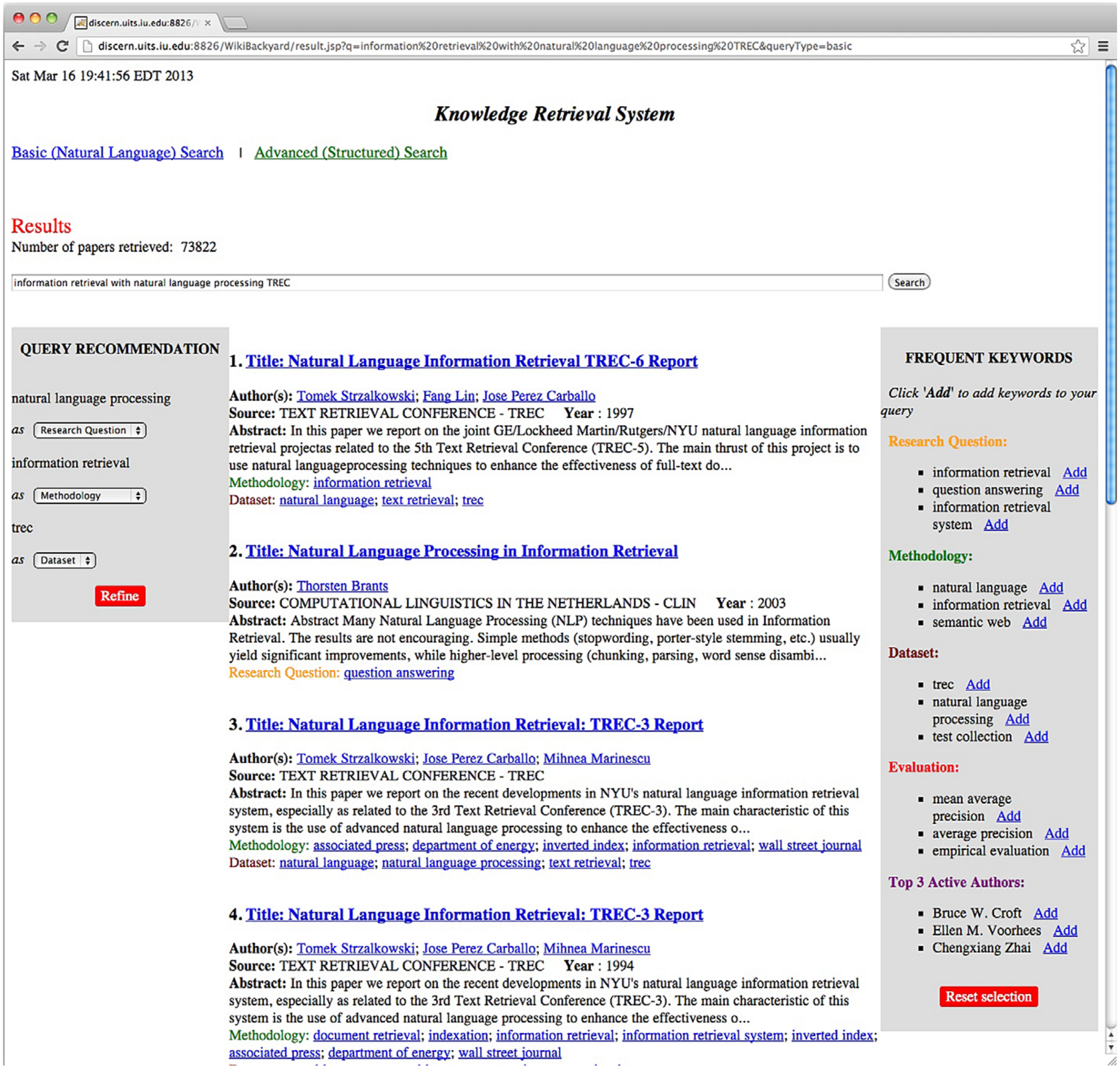

FIG. 1. Knowledge retrieval system (prototype). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

preferences, heuristics, and assumptions), and knowledge context (i.e., understanding the collaboration network). The ability to explicitly identify structural concepts in unstructured text can play an important role in applications such as document summarization (Teufel & Moens, 2002), information retrieval (Tbahriti, Chichester, Lisacek, & Ruch, 2006), information extraction (Mizuta, Korhonen, Mullen, & Collier, 2006), and question answering.

To date, most efforts made to structure knowledge within publications have focused on the medical domain. The Ad Hoc Working Group for Critical Appraisal of the Medical Literature (1987) recommended structural abstracts, as an example, to help individuals assess the content of a publication and to facilitate machine indexing and retrieval. Hirohata, Okazaki, Ananiadou, and Ishizuka (2008) and Lin et al. (2006) automated the process of classifying the sentences of medical publication abstracts as “objectives,” “methods,” “results,” and “conclusions” by using supervised learning and Markov chain analysis. Their classification schema was based on section names found in some scientific abstracts. Similarly, argumentative zoning, an alternative structural schema introduced by Teufel and Moens (2002), can be applied to the abstract or full text of a scientific publication. Following the work of both Teufel and Moens and Mizuta et al. (2006), the following seven categories were adopted by Y. Guo et al. (2010) to explore the structural knowledge within an abstract: background, objective, method, result, conclusion, related work, and future work.

DOI: 10.1002/asi

More recently, Liakata, Teufel, Siddharthan, and Batchelor (2010) introduced a more complex, concept-driven, and ontology-motivated knowledge schema, which employed Core Scientific Concepts, a three-layer annotation scheme treating scientific papers as humanly readable representations of scientific investigations. Eleven concepts are used in the first layer (e.g., “hypothesis,” “motivation,” and “method”). The second layer contains properties of categories (e.g., “advantage,” “disadvantage,” and “method”). The third layer involves coreference identification between the instances of each category. So far, this is one of the most fine-grained and complex structural schemas.

Y. Guo et al. (2010) compared three structural schemas in publication abstracts—section names, argumentative zones, and core scientific concepts (Layer 1)—and found significant relationships and overlap between these three schemas. Section names, actually designed for abstracts, are based on the similar information structure in scientific abstracts (e.g., many abstracts provide the general background and goal of the research, methods used to achieve the goal, the results obtained, or the main conclusions. Not surprisingly, section names, with only four fundamental concepts (objective, method, result, and conclusion), have even sentence distribution within publication abstracts. The sentence distributions of the other two, more complex schemas are skewed, which indicates that structural concepts in the complex schema are missing in a noteworthy percentage of publication abstracts (for instance, “research hypothesis” is not always addressed in the publication abstracts).

From the retrieval point of view, structural knowledge plays an important role in connecting between document-level domain knowledge and the explicit, implicit, or inferred knowledge within user information needs, making it the prerequisite of knowledge retrieval. Lin and Demner-Fushman (2006) proposed a knowledge retrieval framework for the medical domain using knowledge extraction and retrieval methods rooted in the Unified Medical Language System and Medical Subject Headings. Their results showed that knowledge retrieval leveraging structural knowledge is promising when compared with statistical retrieval models in terms of bag-of-word indexing. Hersh et al. (2004) used a “funneling” model to describe the information process for an ideal information retrieval task as: all literature → possibly relevant literature → definitely relevant literature → structured knowledge. However, despite its intuitive appeal, the hypothesis that knowledge retrieval should outperform sophisticated bag-of-word-based retrieval models remains unverified empirically, except in some well-defined domains (e.g., the medical domain), because of the unavailability and sparseness of the knowledge bases. This research develops a framework for automatically generating structural knowledge, which is critical for constructing domain knowledge bases.

Metadata Generation

As structural metadata and some extensions of descriptive metadata are becoming increasingly complex, a remaining problem warrants attention: Who should be responsible for creating and enhancing metadata?

Obviously, for most existing library or document repository systems, professional metadata creators or domain experts (e.g., catalogers and indexers) are the ideal candidates (Milstead & Feldman 1999) to create metadata because they are familiar with the systems and terminologies. However, this approach is costly and may be limited in availability. It is difficult to apply this approach to large amounts of data across different domains.

Meanwhile, other research projects (e.g., Greenberg, Pattuelli, Parsia, & Robertson, 2001) have found that authors can sometimes provide higher quality metadata for web resources; this approach is used by most digital libraries. However, many authors are only willing to provide relatively simple descriptive and reference metadata. Creating more complex metadata (e.g., reasons for citation or structural abstracts) is a demanding job and is not supported comprehensively by most general-domain publications. In the medical domain, some studies have found that explicit structural abstracts are not entirely reliable, and although many core clinical journals require structured abstracts, there is a great deal of variation in the actual headings (e.g., Demner-Fushman & Lin, 2007).

User contributions (i.e., massive numbers of Internet-based volunteers), such as social tagging, are another important resource for metadata generation. According to Bücheler and Sieg (2011), the web offers innovative opportunities for social interaction, collaboration, and collective intelligence. Some popular tagging systems, such as Flickr for image tagging and Delicious for web resource tagging, already have proven to be important and useful metadata-generation sources. Studies on social tagging have found that user-generated metadata tends to be sparse and noisy (e.g., Markines et al., 2009), and this is why social tagging cannot be used to directly generate scholarly metadata.

For the aformentioned reasons, user- (or author-) generated and professional- (or expert-)generated metadata can hardly cope with the need for complex structural metadata generation at a large scale across different domains. Accordingly, the machine-generated approach, an economical and effective alternative, has become popular during the past few years. A wide variety of techniques have been used to process digital texts to generate metadata records. Diekema and Chen (2005) conducted an experiment using NLP and machine learning to assign educational standards to digital content and achieved results comparable to human-created metadata records. Other techniques, such as associative networks (Rodriguez, Bollen, & Van de Sompel, 2009), fuzzy inference (Sah & Wade, 2011), and semiautomatic metadata extraction (Tonkin & Muller, 2008), also have been applied to automatically or semiautomatically creating metadata.

In the medical domain, the ability to accurately model the contextual structure of abstracts is used for MEDLINE publications. McKnight and Srinivasan (2003), for instance, examined the task of categorizing sentences in unstructured medical abstracts using supervised discriminative machine learning techniques. Lin et al. (2006) used a hidden Markov model approach to classify the content of abstracts. Y. Guo et al. (2010) compared the automatic machine learning performance of three different structural schemas for medical abstracts. Teufel and Moens (2002) and Teufel, Siddharthan, and Batchelor (2009) designed a method for automatically identifying the argumentative zones for a publication text.

Most implementations for automatic metadata generation are based on machine learning, which requires a considerable amount of training data. There are two major limitations of this approach. First, although high-quality training structural metadata are available for some domains (e.g., the medical domain), for other domains they are rare. Second, as mentioned earlier, the distribution of structural elements in training texts is quite skewed, and there are only a few training instances for some structural concepts. The number of training instances, however, is critical for machine learning performance (Hirohata et al., 2008).

In this study, to solve this problem, we used a collaborative approach for domain experts and machine and end users via the ScholarWiki system. This approach has proven successful in other fields, for instance, Höök, Rudström, and Waern (1997) used human–machine collaboration to achieve filtered information for hypermedia. Handschuh, Staab, and Ciravegna (2002) implemented semiautomatic annotation of web pages by using S-CREAM, an annotation and authoring framework suited for easy creation of relational metadata.

Because we used human intelligence in this research for scholarly metadata creation, we also propose incentives and an incentive-driven system to help users accomplish this task. This has been identified as an important factor for human-centric systems (Siorpaes & Simperl, 2010).

Structural Metadata Generation

Unlike most existing structural abstract research in the medical domain, we extracted domain-specific structural keywords in our research, based on two assumptions. First, in most existing metadata repositories, author-assigned publication keywords are available as a kind of high-quality descriptive metadata, which can enhance the accuracy of structural keywords; thus, it is not necessary to identify the boundary of keywords as done in named entity recognition (NER) research (i.e., Begin, Inside, Outside) tagging based machine learning. Second, as mentioned earlier, structural keywords are important for implementing knowledge retrieval, that is, for extracting domain knowledge from the user's natural language query (as Figure 1 shows).

For the first assumption, unfortunately, author-assigned keywords are not always readily available or reliable in scientific digital repositories (e.g., ACM or the SciVerse database), but we could employ an automatic approach to enhance the keyword quality. More detailed information can be found in our previous study (C. Guo, Zhang, & Liu, 2013).

In the rest of this section, we describe the artificial intelligence– and human intelligence–based structural keyword generation methods. The experiment in the information retrieval domain is introduced later.

Artificial Intelligence Approach

To harvest comprehensive structural keyword metadata from a large number of publications, we do not use publication full-text as a feature resource because full-text data in most cases are missing or unavailable in scholarly repositories. The three primary sources of features for the structural metadata classifier are the following:

- *Keyword content features:* the nature of the target keyword (i.e., keyword length, keyword content, and capitalization of the keyword).
- *Title context features:* the title context of the target keyword (i.e., the keyword context token and the part of speech [POS] in the publication title).
- *Abstract context features:* the abstract context of the target keyword (i.e., the frequency, location, and context token and the POS of the target keyword in the abstract).

A number of feature types are employed for each category, which can be found in Table 1.

Keyword: *3D model*
Title: Dense sampling and fast encoding for *3D model* retrieval using bag of visual features
Abstract: Our previous shape-based 3D model retrieval algorithm compares 3D shapes by using thousands of local visual features per model. A 3D model is rendered into a set of depth images, and from each image, local visual features are extracted by using the Scale Invariant Feature Transform (SIFT) algorithm by Lowe. To efficiently compare among large sets of local features, the algorithm employs bag-of-features approach to integrate the local features into a feature vector per model. The algorithm outperformed other methods for a data set containing highly articulated yet geometrically simple 3D models. For a data set containing diverse and detailed models, the method did only as well as other methods. This paper proposes an improved algorithm that performs equal or better than our previous method for both articulated and rigid but geometrically detailed models. The proposed algorithm extracts much larger number of local visual features by sampling each depth image densely and randomly. To contain computational cost, the method utilizes GPU for SIFT feature extraction and an efficient randomized decision tree for encoding SIFT features into visual words. Empirical evaluation showed that the proposed method is very fast, yet significantly outperforms our previous method for rigid and geometrically detailed models. For the simple yet articulated models, the performance was virtually unchanged. Categories and Subject Descriptors

DOI: 10.1002/asi

TABLE 1. Feature list for each category.

Category	Feature	Description or example
Keyword Content	Text	Content of the keyword, stemmed, case insensitive, stop words removed • The feature value is "3d model."
	Content_Of_Keyword	A vector of all the tokens in the keyword. • The feature value = ["3d", "model"].
	CAP	Whether the keyword is capitalized • The feature value is "Not All Lower."
	Contain_Digit	If there are digits in the keyword • The feature value is true.
	Character_Length_Of_Keyword	Character length of the target keyword. • The feature value is 2.
	Token_Length_Of_Keyword	No. of tokens in the keyword. • The feature value is 2.
	Category_Length_Of_Keyword	No. of tokens in the keyword; if the length is more than four, we use four to represent its category length • The feature value is 2.
Title Context	Exist_In_Title	Whether keyword exists in title (stemmed, case insensitive, stop words removed) • The feature value is true.
	Location_In_Title	The position the keyword appears in title (not exist, first 3 words, middle, last 3 words) • The feature value is "middle," meaning keyword exists in the middle of the title.
	Title_Text_POS	Unigram and its part of speech in title (in a text window) • The feature values are "for:IN" (−1) and "retriev:NN" (+1).
	Title_Unigram	Unigram of keyword in title (in a text window) . • The feature values are "for" (−1) and "retriev" (+1).
	Title_Bigram	Bigram of keyword in title (in a text window) . • The feature values are "encod for" (−2) and "retriev use" (+2).
Abstract Context	Location_In_Abstract	Which sentence the keyword appears in abstract(first, middle, or last sentence) • The feature value is "first sentence."
	Keyword_Position_In_Sentence_Of_Abstract	Keyword's position in the sentence (beginning, middle, or end) • The feature value is "middle."
	Abstract_Freq	How many times a keyword appears in the abstract • The feature value is 2.
	Abstract_Text_POS	Unigram and its part of speech in abstract (in a text window) • The feature values are "base:VB" (−1) and "retriev:NN" (+1).
	Abstract_Unigram	Unigram of keyword in abstract (in a text window) • The feature values are "base" (−1) and "retriev" (+1).
	Abstract_Bigram	Bigram of keyword in abstract (in a text window) • The feature values are "shape base" (−2) and "retriev algorithm" (+2).

Keyword content feature types, as Table 1 shows, are a feature set characterizing keywords in a context. For instance, CAP characterizes the capitalization nature of the target keyword for a publication, and its value could be one of the following: "All Lower Case" (abc), "All Capitalized" (ABC), "First Char Capitalized" (Abc), or "Not All Lower" (aBc). Content_Of_Keyword is the binary vector typed feature, where each dimension tells whether the corresponding token (in the training vocabulary) appears. To make the content feature more reliable, tokens that occur rarely in the training set are removed. We assume that the length of the keyword is important for classification, and multiple feature types are used to predict the class label (i.e., character length or token length). In the experiment, some keywords contain digits (i.e., "20 newsgroup" or "TREC 2002"), which could be an important clue for some keyword class labels such as "data set." As a result, Contain_Digit is used as a Boolean feature type.

One of the most important contributions of the rhetorical structural metadata is to disambiguate the keyword's role in the publication. For instance, the keyword "clustering algorithm" may be the "research question" or "methodology" for different publications. In this research, we use title context and abstract context features to resolve this problem. Based on NER research experience, a text window [−n, +n] (left *n* words and right *n* words of the target keyword) is used to extract POS, unigram, and bigram features for both title context and abstract context. Instead of using all possible features, we used a cutoff to remove features that occur less than five times, which is effective to remove those noisy features. Location is another important feature type for rhetorical classification, which has been proven as an important feature for argumentative zones research (Merity, Murphy, & Curran, 2009). The numeric title position is used for the Location_In_Title feature; for abstracts, the sentence location (Location_In_Abstract) and keyword position in sentence features (Keyword_Position_In_Sentence_Of_Abstract) are used. For example, the keyword might be observed in the second sentence of the abstract, and it might occur at the beginning of this sentence. If a keyword appears

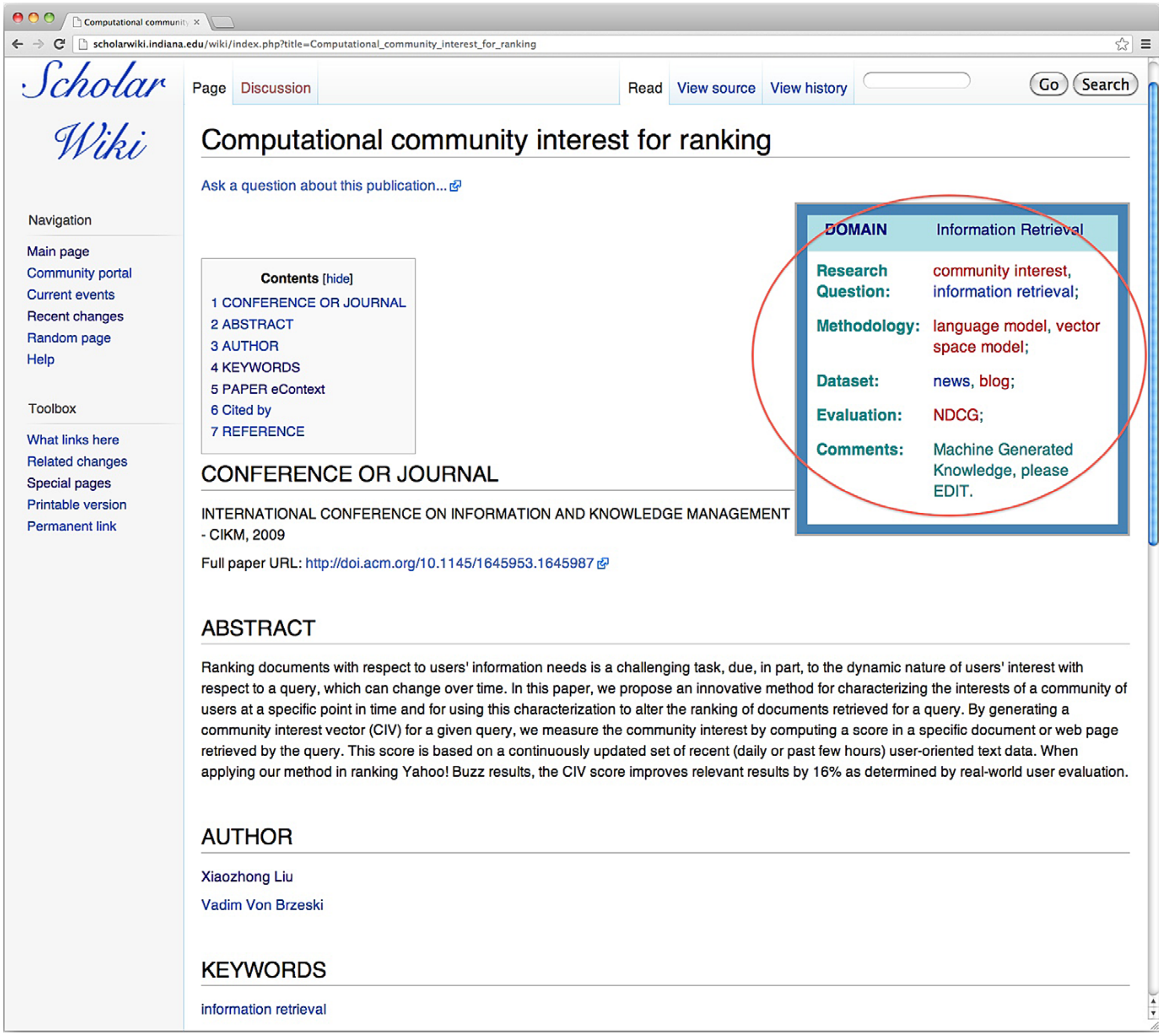

FIG. 2. An example publication ScholarWiki page. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

more than once in the abstract, its first instance is used for feature extraction, and the frequency itself also is a feature (Abstract_Freq).

After comparing different machine learning algorithms, we selected for this study "decision tree," an inductive learning algorithm based on training instances. Decision tree is a simple and effective classification algorithm popular in data mining. It focuses on deducing classification rules represented by a decision tree from a group of random and irregular samples, and we chose the C4.5 implementation of the decision tree algorithm for this study. C4.5, presented by Quinlan (1993), is a successor of ID3, distinguished from the latter by the fact that it selects the attribute with the maximum gain ratio as the splitting attribute (Han, Kamber, & Pei, 2006).

Human Intelligence Approach

Although machine learning and NLP can cope with complex structural metadata generation on a large scale, a fully automated approach tends to suffer from modest precision (discussed later) and often results in noisy or irrelevant results and metadata pollution. In addition, the cost of good-quality training data is high. To solve this problem, we designed the ScholarWiki system.[1] An example publication wiki page is shown in Figure 2. As the example shows, the traditional descriptive metadata—title, author, abstract, citation, and journal information—are presented on the wiki page, and structural keywords, as inferred by a machine learning algorithm, are presented in the top InfoBox.

[1]http://scholarwiki.indiana.edu/wiki/index.php?title=Main_Page

DOI: 10.1002/asi

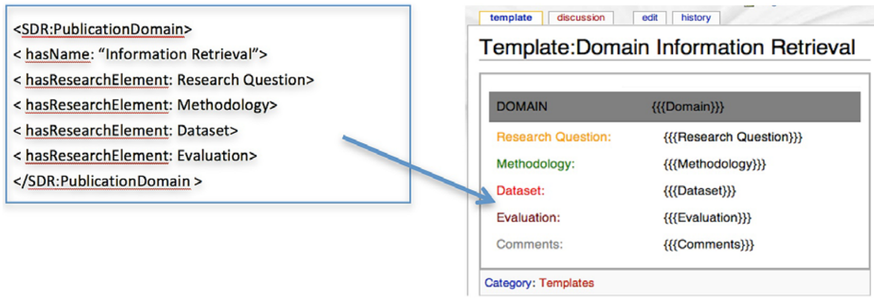

FIG. 3. Domain ontology and ScholarWikiInfoBox template. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

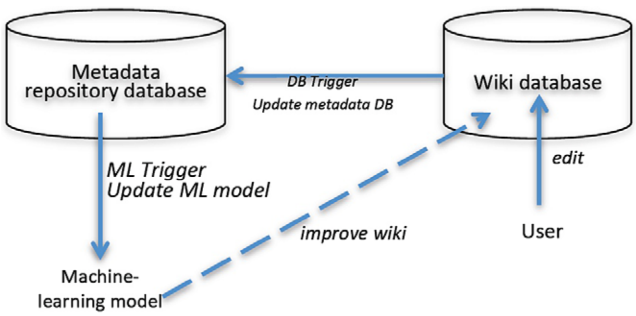

FIG. 4. Two-step "trigger" metadata enhancement. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

By using this system, the domain expert-defined *publication domain schema* (*in RDF format*) can be translated automatically into the ScholarWiki system as a wiki template. For instance, Figure 4 shows how the information retrieval domain RDF is automatically converted into a ScholarWikiInfoBox template.

In the next step, each publication in the target domain is automatically converted into the ScholarWiki style along with the machine-inferred structural keyword information. This progress is implemented in three stages. Initially at the *expert stage*, domain experts provide sample structural keywords via an easy coding system to randomly selected publications in this domain. In the second stage, the *system stage*, a machine learning model with the features described earlier deploys to learn from the domain experts' coding results. The system then infers the structural keywords for all the other publications in this domain and automatically generate wiki pages. All inferred structural metadata are presented in the InfoBox on the top of each publication wiki page. Figure 3 shows a wiki page example generated automatically by the inference algorithm (The circled part is the structural metadata.) Finally, at the *user stage*, users (students, instructors, or scholars) can access and edit the wiki pages to enhance or tweak the metadata presented on the pages if they are not satisfied with the structural keywords or other metadata presented on the wiki page. To facilitate user editing, the ScholarWiki system generates comments on each wiki page. As the following example shows, the system chooses candidate keywords from the publication title or abstract and presents those keywords in the wiki page comments to help users edit the target wiki page.

> *If you need to change the content of the InfoBox, the following candidate keywords may be helpful: [[language model]], [[ranking model]], [[system evaluation]]*

Collaboration Approach

In this metadata-creation method, the user is a central metadata editor instead of just a viewer. We initialize the metadata-generation process by leveraging domain-expert structural ontology definitions and examples of publication-metadata creation. Then, machine-learning algorithms assume control by "understanding" the experts' judgments and automatically generating all the structural metadata for publications in this target domain. Note that even though machines can automatically generate metadata for all publications, the accuracy at this point may not be high given the small amount of training data. In this collaborative process, user editing contributes to (a) improvement in metadata quality by providing judgments on the appropriateness and accuracy of the machine-generated resources (e.g., users can correct errors generated by the machine learning algorithm in the InfoBox) and (b) improvement in machine learning accuracy by adding more training instances. As Figure 5 shows, editing (even if only a small number of publications) triggers (machine-learning triggers) provide comprehensive improvement of structural metadata in the repository database in the target domain. The enhanced repository database can be used to update the quality of all wiki pages in the ScholarWiki system and is thus a nonlinear process, which means that if the user creates or edits his or her publication's metadata, this can trigger the comprehensive improvement of all other publications' metadata. The quality of the metadata, machine-learning model, and wiki pages can be enhanced recursively, which differs from most existing social tagging systems (i.e., Flickr and Delicious).

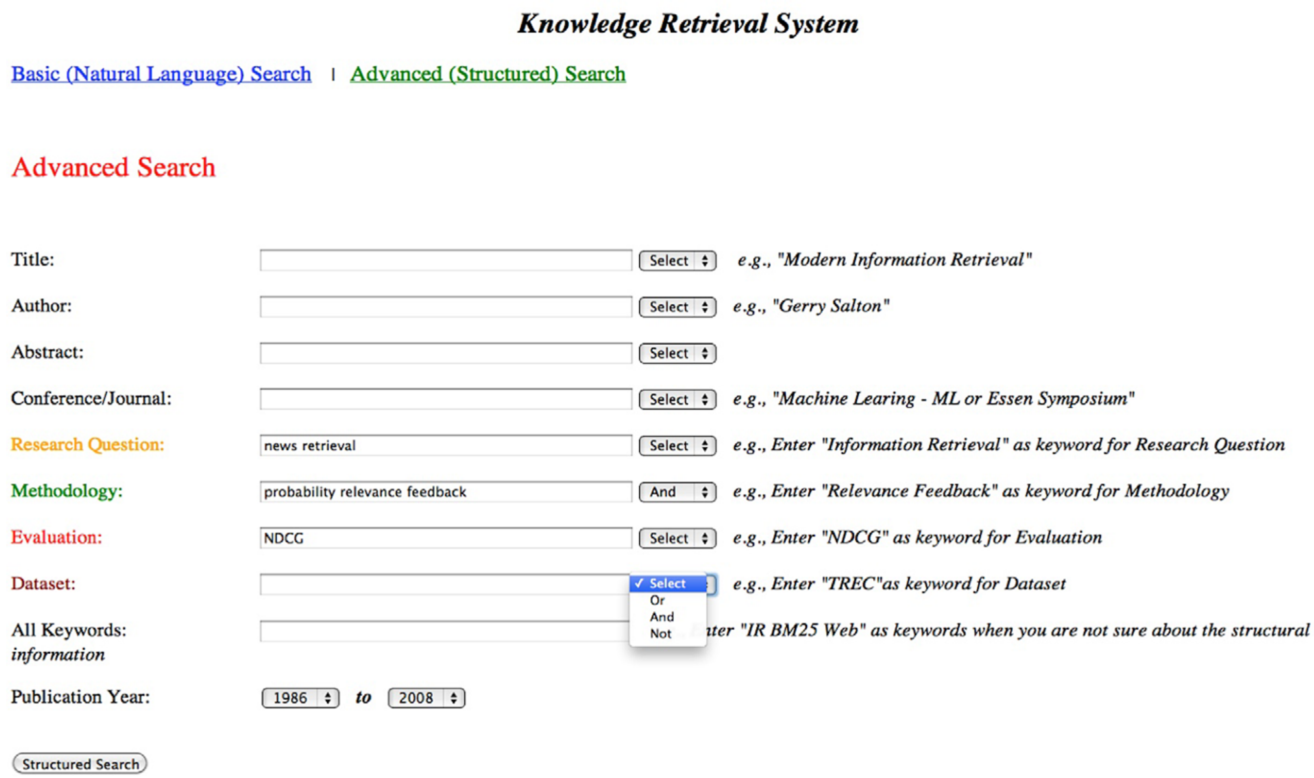

FIG. 5. Knowledge retrieval system prototype. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Incentives

Incentivizing is another essential problem for wiki-oriented social media systems. During the past few years, Wikipedia has successfully attracted millions of users while many similar systems have failed to do so even though they employed similar wiki interfaces. To avoid this kind of problem, the ScholarWiki system prioritizes metadata-presentation rather than data-collection functionality; namely, we want to credit users first, by providing useful metadata and information before collecting data from them. To achieve this goal, we put a question-and-answer (QA) link and eContext metadata (Liu, 2013) on each publication wiki page as referential metadata to accompany the high-quality descriptive metadata. The QA link directs users to a page where they can ask questions (Liu & Jia, 2013) and browse existing questions and answers related to the publication. eContext metadata allow users to get contextual information (e.g., presentation slides or source codes) for the target publication. In addition, to facilitate user browsing, the ScholarWiki system automatically generates a list of author wiki and keyword wiki pages for quick access to resources pertaining to an author or a keyword. Each author wiki page presents the topics and publications that the author has written; the topics of these publications are automatically ranked by citation counts. Each keyword wiki page indexes the most notable authors, the most cited publications, and keywords related to these publications. These features offer a better user experience while providing incentives to use our system (Liu & Qin, in press).

Experiment

To test and compare different approaches such as supervised learning, semisupervised learning, human knowledge-based collaborative learning, and structural keyword generation, we designed an experiment for information retrieval publications. As the collaborative learning approach requires user participation, we invited students from a graduate-level class on information retrieval as users for this experiment.

Data set

In this experiment, 20,799 publications from 1965 to 2010 in the information retrieval domain were used. All the publications came from the *ACM Digital Library*. We first used purposive sampling to identify 15 core conference proceedings and journals in the information retrieval domain (e.g., Special Interest Group on Information Retrieval [SIGIR], Transactions on Information Systems [TOIS], or Conference on Information and Knowledge Management [CIKM]). Publications in these proceedings and journals were used as the seed publications. Cited publications in these were then investigated to expand the

DOI: 10.1002/asi

TABLE 2. Ten-fold cross-validation results on different feature categories (supervised learning).

Feature types	Concept	Precision	Recall	F1
Keyword-based features	Research question	0.549	0.757	0.637
	Methodology	0.727	0.357	0.479
	Data set	0.925	0.742	0.824
	Evaluation	0.800	0.444	0.571
	Weighted average	0.691	0.585	0.603
Keyword + Title-based features	Research question	0.553	0.740	0.633
	Methodology	0.738	0.376	0.498
	Data set	0.925	0.742	0.824
	Evaluation	0.800	0.444	0.571
	Weighted average	0.698	0.584	0.608
Keyword + Title + Abstract-based features	Research question	0.533	0.805	0.642
	Methodology	0.742	0.293	0.420
	Data set	0.942	0.742	0.831
	Evaluation	0.818	0.500	0.621
	Weighted average	0.692	0.581	0.584

corpus. If a paper was cited more than twice by these seed publications, we put them into the test collection. The test collection included 50,617 publications.

In the metadata repository, some publications do not have keyword metadata. To solve this problem, we first created a domain popular keyword (frequency >3) list from the existing keyword collection (for the test collection). Then, we searched each keyword from the paper title and abstract by using greedy matching. (All the titles and abstracts came from *ACM Digital Library*.) For example, if "music information retrieval" was in the title, we would not use the keyword information retrieval. The matched keywords were used as "pseudo keyword" metadata for the target publication if author-assigned keyword metadata were unavailable.

Artificial Intelligence (Machine Learning) Approach Result

Two graduate students with information retrieval backgrounds were hired to annotate keywords for 150 randomly sampled publications from SIGIR and CIKM conference proceedings. Each coder independently annotated 75 papers. The annotated metadata were used for the initial machine learning model training and for evaluation purposes. Before the coders started working on those publications, we trained them for 3 weeks in face-to-face meetings and with coding practice. After training, the coders achieved an 80% agreement rate. The coding process lasted for 4 weeks. If the coders could not decide which category a keyword belonged to in a paper context, they could annotate this keyword as "unknown." If a coder did not have confidence in the target publication (e.g., she did not understand the essence of the paper), she could report and ignore it. Finally, coders annotated 572 paper–keyword pairs for 121 distinct publications. The keyword categories were not evenly distributed; there were 170 research question keywords, 157 methodologies, 62 data set, 18 evaluations, and 165 unknowns.

Based on the feature set presented earlier, we used C4.5 to train the learning model. The text window size of both title and abstract was 3, which means we could use the left 3 words and right 3 words for POS, unigram, and bigram feature extraction in both title and abstract. Table 2 compares 10-fold cross-validation performance of the three basic feature sets: Keyword, Keyword + Title, and Keyword + Title + Abstract.

Given the small amount of training instances, we also implemented a semisupervised learning model with the same feature set and learning algorithm. We selected the top-1% keyword–paper pairs with the highest probability scores for each category in the 15 core journals and conference proceedings. We also used quality control to filter out some low-quality metadata (e.g., we filtered out papers with title length of <50 characters, paper abstract length of <300 characters, and number of keywords ([for the paper] of <5). Finally, 626 keyword–publication pairs, among which there were 200 research question keywords, 200 methodologies, 100 data set, 26 evaluations, and 100 unknowns, were used to expand the original training data set. The performance of semisupervised learning is reported in Table 3.

The detailed result analysis, significance test, and comparison are reported later.

Collaborative Approach Result

Next, we designed and implemented the ScholarWiki system. The supervised learning approach was used to infer structural keyword labels for 20,799 publications in the information retrieval domain. The system automatically generated a wiki page for each publication, with structural keywords presented in the InfoBox section as demonstrated in Figure 2.

Eight graduate students in the information retrieval course were trained to use the system and edit the wiki page, especially for the InfoBox section. For more than 6 weeks, students were asked to use the ScholarWiki system for

TABLE 3. Ten-fold cross-validation results on different feature categories (semisupervised learning).

Feature types	Concept	Precision	Recall	F1
Keyword-based features	Research question	0.580	0.769	0.662
	Methodology	0.724	0.401	0.516
	Data set	0.864	0.773	0.816
	Evaluation	0.800	0.444	0.571
	Weighted average	0.690	0.615	0.627
Keyword + Title-based features	Research question	0.588	0.769	0.667
	Methodology	0.733	0.420	0.534
	Data set	0.864	0.773	0.816
	Evaluation	0.800	0.444	0.571
	Weighted average	0.697	0.622	0.636
Keyword + Title + Abstract-based features	Research question	0.590	0.757	0.663
	Methodology	0.723	0.433	0.542
	Data set	0.879	0.773	0.823
	Evaluation	0.900	0.500	0.643
	Weighted average	0.701	0.624	0.642

TABLE 4. Ten-fold cross-validation results on different feature categories (collaborative approach).

Feature types	Concept	Precision	Recall	F1
Keyword-based features	Research question	0.575	0.793	0.667
	Methodology	0.711	0.408	0.518
	Data set	0.963	0.788	0.867
	Evaluation	0.833	0.556	0.667
	Weighted average	0.701	0.636	0.643
Keyword + Title-based features	Research question	0.579	0.805	0.673
	Methodology	0.733	0.401	0.519
	Data set	0.929	0.788	0.852
	Evaluation	0.833	0.556	0.667
	Weighted average	0.705	0.639	0.644
Keyword + Title + Abstract-based features	Research question	0.576	0.740	0.648
	Methodology	0.639	0.439	0.521
	Data set	0.943	0.758	0.840
	Evaluation	0.800	0.444	0.571
	Weighted average	0.668	0.616	0.627

class-reading downloads, group-assignment implementations, class-presentation preparation, and final-project resource retrieval. Students were encouraged to edit the wiki pages if they found mistakes or wanted to recommend additional information in the structural keyword part. The class instructor also could edit the wiki pages, if necessary.

The system triggers, at the back end, automatically extracted new training instances from the users' editing behavior. The machine-learning model was enhanced periodically based on the new instances. At the end of the semester, the updated machine learning model was used for this article. Its performance is shown in Table 4.

For the entire semester, students and the instructor edited 205 keyword–publication pairs with 92 research question keywords, 67 methodologies, 27 data set, 17 evaluations, and 2 unknowns belonging to 57 distinct publications. Note that the number of user-contributed training instances for a collaborative approach is much smaller than those for the semisupervised learning approach. Similar to other wiki systems, the wiki page crowd-sourced approach renders editing a collaborative progress wherein multiple people can work on the same wiki page.

Analysis

In this experiment, we compared three different approaches to structural keyword metadata-creation: supervised learning, semisupervised learning, and (user + machine) collaborative learning with human intelligence. To compare their performance, the F1 scores of these different methods with different feature sets are presented in Table 5. We also implemented a significance test (*t* test) using the supervised learning approach as the baseline algorithm. Note that even though the training instances are different for these approaches, the testing set of these methods—the domain experts' coding results—is the same. Specifically, while conducting 10-fold cross-validation, for supervised learning, we

DOI: 10.1002/asi

TABLE 5. F measure comparison for supervised learning, semisupervised learning, and collaborative learning (with significance test using supervised learning as baseline).

F1 compare	Concept	Supervised	Semisupervised	Collaborative
Keyword-based features	Research question	0.637	0.662	**0.667*** **
	Methodology	0.479	0.516**	**0.518****
	Data set	0.824	0.816	**0.867***
	Evaluation	0.571	0.571	**0.667***
Keyword + Title-based features	Research question	0.633	0.667	**0.673*** **
	Methodology	0.498	**0.534****	0.519**
	Data set	0.824	0.816	**0.852***
	Evaluation	0.571	0.571	**0.667***
Keyword + Title + Abstract-based features	Research question	0.642	**0.663****	0.648**
	Methodology	0.420	**0.542*** **	0.521**
	Data set	0.831	0.823	**0.840**
	Evaluation	0.621	**0.662****	0.571**

*$p < .01$; **$p < .005$; ***$p < .001$.

divided the domain experts' coding results into 10 parts and took nine parts as training instances and one part as testing instances. For semisupervised learning, the domain experts' coding results and the instances inferred from the supervised learning were divided into 10 parts, and we took nine parts of them as training instances and one part of the coding results as testing instances. For collaborative learning, the training and testing instances were obtained in a similar way as those for semisupervised learning. In this experiment, we assume that the coding results in this research are reliable. For each row, the best performance approach is marked in bold.

Because it is difficult to estimate if the experiment results follow a normal distribution, we also used the Wilcoxon signed-rank test, a nonparametric statistical-hypothesis test, to validate the hypothesis. Results show that the collaborative approach is significantly better than is the supervised approach for research question (for all kinds of features), methodology (for all kinds of features), and evaluation (Keyword + Title + Abstract-based features).

We summarize our findings thus:

- Overall, the best-performing result was the collaborative learning approach with Keyword + Title features. Similarly, collaborative learning achieved the best performance with Keyword features.
- Both semisupervised learning and collaborative learning effectively (significantly) improved the accuracy of the supervised learning method. Overall, the collaborative approach outperformed the semi-learning approach. Given the short period (6 weeks) of students' participation, it is difficult to assume the trade-off between the amount of human time spent and the accuracy enhancement, and this should be explored in future studies.
- For feature set validation, although the title context feature effectively optimized the performance for all learning approaches, the abstract context features could not offer stable improvement for collaborative learning approaches. However, as abstract context features significantly enhanced supervised learning except for the methodology item, and semisupervised learning performance except for the research question item.
- Data set items were the most accurately predicted, and collaborative learning was effective at further enhancing performance. In comparison, methodology items presented the most difficult task, and semisupervised learning was the most effective method.

In our initial hypothesis, both title and abstract context features were thought to be important for enhancing machine-learning performance. In the experiment, however, we found that abstract features only help the semisupervised learning approach. The reason for this might be the number of training instances. Given the small number of training instances, when we extract a [−3, +3] text window for unigram, POS, and bigram features, abstract context features cannot make a great contribution to the learning performance due to their large variation. Semisupervised learning, on the contrary, significantly expanded the training set with 626 new instances. This might be the main reason why abstract context features work well for this approach, especially for method items, which highly depend on the context features.

Compared with the semisupervised learning approach, the collaborative learning approach, with only 205 additional training instances, did not have a chance to improve the abstract context features. However, the Keyword + Title context features for this approach achieved the best performance in this experiment, which proves that the quality of human intelligence-generated structural keyword instances is much higher than is that of the machine-generated instances. On the other hand, for the data set and evaluation labels, coders annotated relatively few training instances (66 and 18, respectively). Small numbers of training instances negatively affect the performance of semisupervised learning, especially for the keyword content-based features, in which a machine may straightforwardly memorize the nature of those keywords (e.g., the length or number of keyword tokens). A human intelligence–based collaborative learning approach successfully solves this problem because user-contributed metadata via ScholarWiki system are not necessarily restricted or related to the initial machine learning results. This is probably why the accuracy for the data set and evaluation labels is significantly higher for the collaborative learning approach.

Finally, we found that the keyword-labeling method does not perform as well as others. There may be two reasons for this. First, coders and ScholarWiki system users may occasionally have confusion between research question and methodology for some publications. For example, while some students may have annotated "language model" as the research question (for a paper), others may have tagged it as "methodology." Second, research question and methodology context features sometimes may be confusing. For example, research question can appear at the beginning of a title, but authors sometimes may start the title with methodology keywords. To solve this problem in the future, we will use a more sophisticated feature set (e.g., sentence tense, parsing tree, or semantic role labeling features) to better disambiguate research question and methodology keywords in a rich context. Another possibility is to explore full-text features such as section names, section numbers, or keyword-occurrence distribution, even though we deliberately avoided the use of full text in this study because of access difficulty. For example, research question keywords frequently appear in the first section of the article whereas methodology-related keywords usually occur in the third or fourth section of the article.

Structural Keyword Use Cases

In this section, we provide a couple of use cases of structural keywords, which demonstrate the usefulness of this innovative structural metadata.

Knowledge-Based Information Retrieval

The success of several knowledge-retrieval experiments in some domains (e.g., Lin & Demner-Fushman, 2006) proved empirically that knowledge retrieval, if implemented in the restricted domain, could yield performance gains over classical bag-of-word-based retrieval systems. The key to their success is the availability of codified domain knowledge accumulated in years. However, this knowledge base is rarely available in other domains. Structural keyword metadata along with collaborative learning method we propose in this article enable the implementation of a prototype knowledge-retrieval system to serve users' knowledge-oriented information needs. The system could benefit users by providing structure-level access to scientific papers and automatically inferring structural knowledge from their natural language queries by leveraging knowledge recommendation and knowledge feedback (as Figure 1 shows). Similarly, users also can provide explicit structural queries by using the interface provided in Figure 5.[2]

Unlike the traditional "bag-of-word" assumption, each publication in the knowledge-retrieval system is indexed with "bag-of-knowledge" representations, and different domain knowledge is attached to the same keyword for different publications, which can be used to further understand user information need.

Knowledge-Based Bibliometrics

Scientific publications with scientific metadata can be represented in various types of forms for in-depth analysis. Domain knowledge graph modeling is one of the most important ways to describe the significant characters of a selected domain, which is essential for domain information visualization, information retrieval, bibliometric analysis, and scientific literature discovery. The constructed domain knowledge graph represents not only the important scholar components (i.e., authors, keywords, publications, or journals) of the domain but also the illustration of the social network relatedness between them (i.e., publication relatedness, coauthorship, or keyword co-occurrence probability).

Keyword and domain knowledge provide a new opportunity to unveil the unexplored information by leveraging classical methods in bibliometrics. Taking the information retrieval domain as an example, a keyword relationship graph is constructed by using keyword co-occurrence probability (of 50,617 publications) in Figure 6. Unlike the classical keyword co-occurrence graph presented by Lee, Su, and Chan (2010) and Kajikawa and Takeda (2008), this is a heterogeneous graph, and each vertex type (associated with a color) represents a specific knowledge type (i.e., research question, methodology, data set, and evaluation). If key_i as a research question and key_j as a methodology are connected on the graph, key_j often provides a methodological solution for key_i in the domain.

By using this graph, we can further investigate the following exemplar questions for the information retrieval domain:

- What is the most popular research question/methodology/data set/evaluation method in the information retrieval domain?
- Which methodology can provide solution for what kind of research question(s)?
- By using a community identification algorithm, for each community on the heterogeneous graph, what is the methodological hub, and what is the data set/evaluation hub?

Conclusion

In this article, we proposed three different approaches to generating a large number of domain-specific, novel

[2]The prototype retrieval system: http://discern.uits.iu.edu:8826/WikiBackyard/StructuredSearch.jsp

DOI: 10.1002/asi

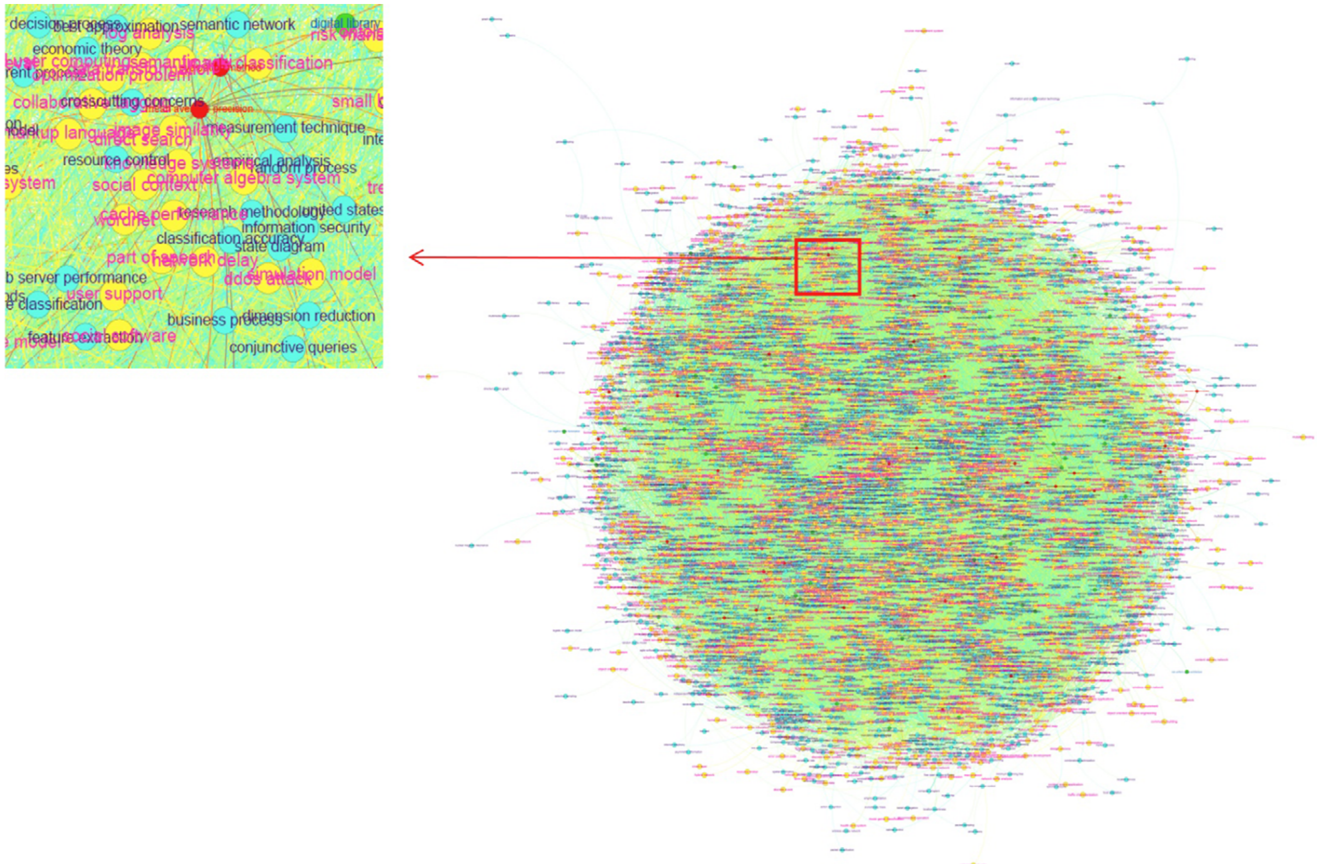

FIG. 6. Heterogeneous graph with four types of vertices (each color represents a specific knowledge type). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

structural metadata—structural keywords—for scientific publications (for 50,617 publications in the information retrieval domain). By using artificial intelligence methods (supervised learning and semisupervised learning), machine learning can automatically generate structural keywords at a low cost.

The other contribution of this article is to introduce a user-enhanced, machine-learning scientific metadata-creation method. By allowing low-barrier participation, system users (instructors, scholars, and students) can participate in the knowledge and metadata editing and enhancing process. Unlike other professional metadata-management systems in which users need considerable expertise to contribute, ScholarWiki users do not need to understand the detailed metadata schema. Their job is to simply edit the wiki page, especially the InfoBox section, for structural keyword enhancement.

ScholarWiki differs from other wiki-based social media systems in three important aspects. First, all the wiki pages along with structural keyword metadata are initially generated by the system via artificial intelligence. Second, a user's contribution to ScholarWiki is nonlinear and evolving; in effect, editing one wiki page triggers the comprehensive improvement of all other wiki pages. Third, although users can learn scientific keywords and publications from ScholarWiki, the system is simultaneously learning from users by "understanding" their editing behaviors and practices. In other words, the system trigger, at the back end, is the platform for bidirectional metadata communication (from the system to users, and vice versa).

Our experiment in the information retrieval domain with a graduate-level class proves that the human + artificial intelligence approach (collaborative learning) is an effective and economical way to create innovative, knowledge-based metadata. This method can be generalized to other academic domains at low cost, if the target domain specific structure and a small amount of the training data set are available.

References

Ad Hoc Working Group for Critical Appraisal of the Medical Literature. (1987). A proposal for more informative abstracts of clinical articles. Annals of Internal Medicine, 106(4), 598–604. doi: 10.7326/0003-4819-106-4-598

Allen, B., Qin, J., & Lancaster, F.W. (1994). Persuasive communities: A longitudinal analysis of references in the philosophical transactions of the Royal Society, 1665-1990. Social Studies of Science, 24(2), 279–310. doi: 10.1177/030631279402400204

Bücheler, T., & Sieg, J.H. (2011). Understanding science 2.0: Crowdsourcing and open innovation in the scientific method. In Proceedings of the Second European Future Technologies Conference and Exhibition (pp. 327–329).

Demner-Fushman, D., & Lin, J. (2007). Answering clinical questions with knowledge-based and statistical techniques. Computational Linguistics, 33(1), 63–103.

Diekema, A.R., & Chen, J. (2005). Experimenting with the automatic assignment of educational standards to digital library content. In Proceedings of the Fifth ACM/IEEE-CS Joint Conference on Digital Libraries (pp. 223–224). New York, NY: ACM.

Evans, J.A., & Foster, J.G. (2011). Metaknowledge. Science, 331(6018), 721–725.

Greenberg, J., Pattuelli, M., Parsia, B.W., & Robertson, D. (2001). Author-generated Dublin Core metadata for web resources: A baseline study in an organization. In K. Oyama & H. Gotoda (Eds.), Proceedings of the International Conference on Dublin Core and Metadata Applications (pp. 38–46).

Guo, C., Chinchankar, R., & Liu, X. (2012). Knowledge retrieval for scientific literatures. Proceedings of the American Society for Information Science and Technology, 49(1), 1–7. doi: 10.1002/meet.14504901152

Guo, C., Zhang, J., & Liu, X. (2013). Scientific metadata quality enhancement for scholarly publications. iConference 2013 Proceedings (pp. 777–780). doi:10.9776/13382

Guo, Y., Korhonen, A., Liakata, M., Karolinska, I., Sun, L., & Stenius, U. (2010). Identifying the information structure of scientific abstracts: An investigation of three different schemes. In Proceedings of the Workshop on Biomedical Natural Language Processing (pp. 99–107).

Han, J., Kamber, M., & Pei, J. (2006). Data mining: Concepts and techniques (2nd ed.). San Francisco, CA: Kaufmann.

Handschuh, S., Staab, S., & Ciravegna, F. (2002). S-CREAM Semi-automatic CREAtion of Metadata. In A. Gómez-Pérez & V.R. Benjamins (Eds.), Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management (pp. 358–372).

Hersh, W., Cohen, A., Yang, J., Bhupatiraju, R.T., Roberts, P., & Hearst, M. (2004). TREC 2005 Genomics Track overview. In Proceedings of the 14th Text REtrieval Conference (TREC 2005). Retrieved from http://trec.nist.gov/pubs/trec14/papers/GEO.OVERVIEW.pdf

Hirohata, K., Okazaki, N., Ananiadou, S., & Ishizuka, M. (2008). Identifying sections in scientific abstracts using conditional random fields. In Proceedings of the Third International Joint Conference on Natural Language Processing (pp. 381–388).

Höök, K., Rudström, A., & Waern, A. (1997, April). Edited adaptive hypermedia: Combining human and machine intelligence to achieve filtered information. Paper presented at the Flexible Hypertext Workshop at the Eighth Association for Computing Machinery (ACM) International Hypertext Conference (Hypertext'97). Retrieved from http://soda.swedish-ict.se/2602/1/Hypertext97.html

International Federation of Library Associations. (1998). Functional requirements for bibliographic records: Final report. Munich, Germany: K.G. Saur Verlag. Retrieved from http://archive.ifla.org/VII/s13/frbr/frbr.pdf

Kajikawa, Y., & Takeda, Y. (2008). Structure of research on biomass and bio-fuels: A citation-based approach. Technological Forecasting and Social Change, 75(9), 1349–1359. doi:10.1016/j.techfore.2008.04.007

Lee, P.C., Su, H.N., & Chan, T.Y. (2010). Assessment of ontology-based knowledge network formation by vector-space model. Scientometrics, 85(3), 689–703. doi:10.1007/s11192-010-0267-8

Liakata, M., Teufel, S., Siddharthan, A., & Batchelor, C. (2010, May). Corpora for the conceptualisation and zoning of scientific papers. Paper presented at the Seventh International Conference on Language Resources and Evaluation.

Lin, J., & Demner-Fushman, D. (2006). The role of knowledge in conceptual retrieval: A study in the domain of clinical medicine. In Proceedings of the 29th annual International Association for Computing Machinery Special Interest Group on Information Retrieval Conference on Research and Development in Information Retrieval (ACM SIGIR) (pp. 99–106).

Lin, J., Karakos, D., Demner-Fushman, D., & Khudanpur, S. (2006). Generative content models for structural analysis of medical abstracts. In Proceedings of the Workshop on Linking Natural Language Processing and Biology: Towards Deeper Biological Literature Analysis (pp. 65–72).

Liu, X. (2013). Generating metadata for cyberlearning resources through information retrieval and meta-search. Journal of the American Society for Information Science and Technology, 64(4), 771–786. doi: 10.1002/asi.22744

Liu, X., & Jia, H. (2013). Answering academic questions for education by recommending cyberlearning resources. Journal of the American Society for Information Science and Technology doi: 10.1002/asi.22851.

Liu, X., & Qin, J. (in press). An interactive metadata model for structural, descriptive, and referential representation of scholarly output. Journal of the American Society for Information Science and Technology.

Markines, B., Cattuto, C., Menczer, F., Benz, D., Hotho, A., & Stumme, G. (2009). Evaluating similarity measures for emergent semantics of social tagging. In Proceedings of the 18th International Conference on World Wide Web (pp. 641–650).

McKnight, L., & Srinivasan, P. (2003). Categorization of sentence types in medical abstracts. In Proceedings of the American Medical Informatics Association (AMIA) 2003 Annual Symposium (pp. 440–444).

Merity, S., Murphy, T., & Curran, J.R. (2009). Accurate argumentative zoning with maximum entropy models. In Proceedings of the Workshop on Text and Citation Analysis for Scholarly Digital Libraries (pp. 19–26).

Milstead, J., & Feldman, S. (1999). Metadata: Cataloging by any other name. Online, 23(1), 24–31.

Mizuta, Y., Korhonen, A., Mullen, T., & Collier, N. (2006). Zone analysis in biology articles as a basis for information extraction. International Journal of Medical Informatics, 75(6), 468–487.

Quinlan, J.R. (1993). C4.5: Programs for machine learning. San Francisco, CA: Kaufmann.

Rodriguez, M.A., Bollen, J., & Van de Sompel, H. (2009). Automatic metadata generation using associative networks. ACM Transactions on Information Systems, 27(2), 1–20.

Sah, M., & Wade, V. (2011). Automatic mining of cognitive metadata using fuzzy inference. In P. De Bra & K. Grønbæk (Eds.), Proceedings of the 22nd Association for Computing Machinery Conference on Hypertext and Hypermedia (pp. 37–46). New York, NY: ACM.

Shotton, D., Portwin, K., Klyne, G., & Miles, A. (2009). Adventures in semantic publishing: Exemplar semantic enhancements of a research article. PLoSComputional Biology, 5(4), e1000361. doi:10.1371/journal.pcbi.1000361

Siorpaes, K., & Simperl, E. (2010). Human intelligence in the process of semantic content creation. World Wide Web Journal, 13(1), 33–59.

Tbahriti, I., Chichester, C., Lisacek, F., & Ruch, P. (2006). Using argumentation to retrieve articles with similar citations: An inquiry into improving related articles search in the MEDLINE digital library. International Journal of Medical Informatics, 75(6), 488–495.

Teufel, S., & Moens, M. (2002). Summarizing scientific articles: Experiments with relevance and rhetorical status. Computational Linguistics, 28, 409–445.

Teufel, S., Siddharthan, A., & Batchelor, C.R. (2009). Towards domain-independent argumentative zoning: Evidence from chemistry and computational linguistics. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (pp. 1493–1502).

Tonkin, E., & Muller, H.L. (2008). Semi-automated metadata extraction for preprints archives. In R. Larsen, A. Paepcke, J. Borbinha, & M. Naaman (Eds.) Proceedings of the Eighth ACM/IEEE-CS Joint Conference on Digital Libraries (pp. 157–166). New York, NY: ACM.

Zeng, M.L., & Qin, J. (2008). Metadata. New York, NY: Neal-Schuman.