

# Chapter 9

## Searching and Generating Authoring Information: A Hybrid Approach

**Faisal Alkhateeb**

*Yarmouk University, Jordan*

**Iyad Abu Doush**

*Yarmouk University, Jordan*

**Amal Alzubi**

*Yarmouk University, Jordan*

**Shadi Aljawarneh**

*Isra University, Jordan*

**Eslam Al Maghayreh**

*Yarmouk University, Jordan*

### ABSTRACT

*In this paper, the authors propose a novel approach to search and retrieve authoring information from online authoring databases. The proposed approach combines keywords and semantic-based methods. In this approach, the user can retrieve such information considering some specified keywords and ignore how the internal semantic search is being processed. The keywords entered by the user are internally converted by the system to a semantic query that will be used to search the requested information. The authors then use (X)HTML-based templates for the automatic construction of BibTeX elements from the query results.*

### 1. INTRODUCTION

The World Wide Web (WWW) has become the first source of knowledge for many subjects. It can be seen as an extensive information system that allows exchanging the resources as well as documents. The semantic web is an evolving ex-

tension of the web aiming at giving well defined forms and semantics to the web resources (e.g., content of an HTML web page) (Berners-Lee et al., 2001).

Due to the growth of the semantic web, semantic search has become an attracting area. The term refers to methods of searching web

DOI: 10.4018/978-1-4666-2157-2.ch009

documents beyond the syntactic level of matching keywords. Exposing metadata is an essential point for a semantic search approach associated with the semantic web. RDF (Resource Description Framework) (Manola & Miller, 2004) is a knowledge representation language dedicated to the annotation of resources within the Semantic web. Currently, many documents are annotated via RDF due to its simple data model and its formal semantics. For example, RDF can be embedded in (X)HTML web pages using the RDFa annotations (Adida & Birbeck, 2008) and in SMIL documents (Bulterman et al., 2005) using RDF/XML (Beckett & McBride, 2004). SPARQL (Prud'hommeaux & Seaborne, 2008) is a W3C recommendation language developed in order to query RDF knowledge bases, e.g., retrieving nodes from RDF graphs. In addition, it can be used to construct RDF graphs from the instantiation of the query result to the graph pattern (i.e., an RDF graph with variables) specified in the CONSTRUCT clause of the query.

Semantic web languages (i.e., RDF and OWL) can be used for knowledge encoding and it can be used by services, tools and applications (Finin & Ding, 2006). The semantic web will not help only human to search web contents, but also machines will be able to process web contents. This can help in creating intelligent services, customized web and can have more powerful search engines (Decker et al., 2000).

The common approach for searching web contents is based on using keywords. More precisely, both queries and documents are typically treated at a word or gram level. The search engine is missing a semantic-level understanding of the query and can only understand the content of a document by picking out documents with the most commonly occurring keywords.

The objective of this paper is to provide a novel approach for retrieving authoring information that combines keyword-based and semantic-based approaches. In this approach, the user is interested only in retrieving authoring information considering some specified keywords and ignores how

the internal semantic search is being processed. In particular, the user is interested in searching authoring information from online authoring information portals (such as DBLP, ACM, IEEE, etc.). For example, show me all documents of the author “faisal alkhatieb” or “jerome euzenat” with a title containing “SPARQL.” In the proposed approach, keywords are used for collecting authoring information about the authors, which are then filtered with semantic search (using RDF and SPARQL) based on the semantic relations of the query. After that, the query results are used to generate BibTeX elements using (X)HTML-based templates.

The remainder of the paper is organized as follows: a review of related work is discussed in Section 2. We introduce the research background in Section 3. The proposed approach is presented in Section 4 as well as a case study will be illustrating the proposed approach. Section 5 provides the details on generating BibTeX elements using (X)HTML-based templates. Conclusions drawn from this study are presented in Section 6.

## 2. RELATED WORK

Traditional search engines use keywords as their search basis. Semantic search applies semantic processing on keywords for a better retrieval search. Hybrid search utilizes the keyword search from regular search along with the ability to use semantic search to query and reason using metadata. Using ontologies, the search engines can find pages that have different syntax but similar semantics (Decker et al., 2000).

The hybrid search provided users with more capabilities for searching and reasoning to obtain enhanced results. According to Bhagdev et al. (2008) there are three types of queries using hybrid search:

- Semantic search using the defined metadata and the relations between instances.

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/searching-generating-authoring-information/72758](http://www.igi-global.com/chapter/searching-generating-authoring-information/72758)

## Related Content

---

### **Towards Efficient Big Data Storage With MapReduce Deduplication System**

Vijesh Joe, Jennifer S. Rajand Smys S. (2021). *International Journal of Information Technology and Web Engineering* (pp. 45-57).

[www.irma-international.org/article/towards-efficient-big-data-storage-with-mapreduce-deduplication-system/275733](http://www.irma-international.org/article/towards-efficient-big-data-storage-with-mapreduce-deduplication-system/275733)

### **Machine Learning and Web Mining: Methods and Applications in Societal Benefit Areas**

Georgios Lappas (2010). *Web Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 1495-1514).

[www.irma-international.org/chapter/machine-learning-web-mining/37700](http://www.irma-international.org/chapter/machine-learning-web-mining/37700)

### **Mobile Engagement: Dynamics of Transmedia Pervasive Narratives**

Elizabeth Evans (2016). *Web Design and Development: Concepts, Methodologies, Tools, and Applications* (pp. 734-747).

[www.irma-international.org/chapter/mobile-engagement/137373](http://www.irma-international.org/chapter/mobile-engagement/137373)

### **Towards Improving the Lexicon-Based Approach for Arabic Sentiment Analysis**

Nawaf A. Abdulla, Nizar A. Ahmed, Mohammed A. Shehab, Mahmoud Al-Ayyoub, Mohammed N. Al-Kabiand Saleh Al-rifai (2014). *International Journal of Information Technology and Web Engineering* (pp. 55-71).

[www.irma-international.org/article/towards-improving-the-lexicon-based-approach-for-arabic-sentiment-analysis/123184](http://www.irma-international.org/article/towards-improving-the-lexicon-based-approach-for-arabic-sentiment-analysis/123184)

### **Benefits and Challenges for Business Process Management in the Cloud**

Ute Riemann (2016). *Web-Based Services: Concepts, Methodologies, Tools, and Applications* (pp. 2096-2121).

[www.irma-international.org/chapter/benefits-and-challenges-for-business-process-management-in-the-cloud/140890](http://www.irma-international.org/chapter/benefits-and-challenges-for-business-process-management-in-the-cloud/140890)