

# Semantic Knowledge Management System

Paripati Lohith Kumar

**Abstract**—The scholarly activities query, retrieval, maintenance and management of institution repository data is often difficult, because repositories are loaded with documents in a myriad of formats generated by an Institution. The current data retrieval techniques that are used in institutional repositories make use of indexing and keywords in order to retrieve information which results in reduced accuracy in matching user requests.

This paper proposes a system called SEMKNOW<sup>1</sup>, which will enable machine involvement to enhance the efficiency of information seeking process through flexible query option and well conceptually organized structure of documents inside the repository for researchers.

Adding semantic technologies to the educational repository of the institution will add support for the functions carried out by the various roles involved in knowledge development and dissemination in academic institutions.

**Index Terms**— Ontology, Semantic web, Web 3.0, Institution repository, Knowledge management system.

## 1 INTRODUCTION

ONE of the main objectives of universities is the dissemination and creation of knowledge. It is important for an academic institution to perpetuate and evolve knowledge. To this end they make use of digital repositories in order to easily access, maintain and distribute materials that are developed in the course of conducting research. The nature of distribution of these materials is more user-centric and more focus is placed on creating relevant folksonomies in order to enable easier retrieval and access for the user, however, the retrieval process would be more effective if the tags were machine understandable and so the system would be actively involved in aiding the user to find relevant documents.

The open source SEMKNOW project aims to provide intelligent system that agreements with in-built interfaces for the *representation, integration, management and querying* of knowledge through semantic web. In this paper, I describe an initial web-based application that adopts an ontology centric model to perform semantic query answering over Publications (Conference papers, Theses, E-Journals, E-Books, Tech Reports, Manuals, Unpublished articles, Research and Development Projects). Ontologies in this application are described using OWL 2 (Web Ontology Language).

This paper is organized as follows. The next section presents the problem statement and motivation for my work. It is followed by a section describing the SEMKNOW architecture and implementation. Next I describe the SEMKNOW systems domain ontology, application ontology and how they both interoperate with each other. This is followed by a section that describes how this are represented in RDF (Resource Description Framework). Then I go on to describe the prototype built and finally I present the conclusions of my work.

## 2 MOTIVATION

In the Knowledge-intensive view of education, it requires knowledge bases and knowledge systems built to effectively store and retrieve the generated knowledge in the right context to advance learning. Over recent years universities and higher education institutions are pioneering knowledge management into their organization and there is quite tremendous growth in publications by researchers, contributing to population of web with research literature.

A free platform for knowledge sharing not only among academia, but also for students and other researchers around the globe is seen as the result of digital libraries, Institutional repositories emergence. Nevertheless, the complex collaboration processes that take place in the teaching-learning process still poses major challenges to research mainly in Knowledge Management itself and Artificial Intelligence as a supporting field due to the informality that in most cases surrounds learning.

Today technologies, initiatives and strategies such as the OWL, RDF, and DCMI (Dublin Core Metadata Initiative) [1] subsist or are being developed which allow identifying and describing knowledge and information resources. It is therefore important that knowledge management technologies and strategies be researched, developed and applied in education to practise the future workforce for the new economic model and simultaneously enrich their learning environment.

It is for that reason so essential to develop ontologies and tools that allow managing different types of documents that build research literature and knowledge needs to be stored and retrieved in right context. This stated necessity generates a justifiable motivation for the development of tools to assist students and professors in storing and retrieving information documents that result from their teaching-learning process

• Paripati Lohith Kumar is currently pursuing MS degree program in software engineering in Vellore Institute of Technology University, India, PH-+919597737793. E-mail: plohitkumar@hotmail.com

<sup>1</sup> SEMKNOW – Semantic Knowledge management System

and thus generate knowledge-centric collaboration beyond the confines of classroom.

### 3 SYSTEM ARCHITECTURE

SEMKNOW was implemented using the Java programming language and interoperates with other open source and Semantic Web technologies [2], including the Sesame<sup>2</sup> Java API, Pellet<sup>3</sup> and Protégé<sup>4</sup>. The system architecture shown in Fig.1 is composed of several components including an application aggregator, a web user interface, an ontology repository, an ontology index, a reasoner that exhibit the inferred relations, properties, classes and a SeRQL (Sesame RDF Query Language) query engine for querying a triple store [3].

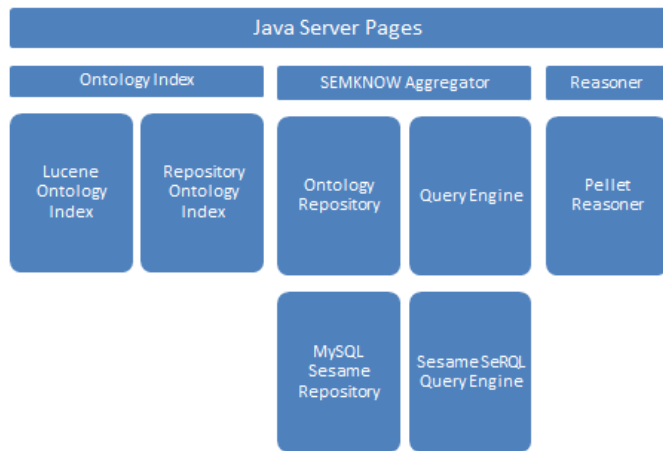


Fig. 1. SEMKNOW Architecture

#### 3.1 SEMKNOW Aggregator

The SEMKNOW aggregator is the nucleus element of the system which integrates the decoupled relations among systems modules, operates the system behavior, and manages ontologies used by the application. The aggregator loads requested ontology as per the context and the retrieved ontology. The aggregator enables communication between components. The manipulation of OWL documents and data structures by multiple SEMKNOW components is facilitated by the Sesame Java API.

#### 3.2 Ontology Repository

Storage and retrieval of URI identified ontologies is supported by ontology repository. This implementation in SEMKNOW system uses a sophisticated MySQL database storage.

#### 3.3 Ontology Indexer

The ontology indexer though behaves like lookup table; it enables fast and easy access to entities in ontology. Indexed entities include class, property, their types and individual

names. Thus, this component supports partial entity name matches, and retrieval of specific types of entities. The current implementation leverages the speed and power of the Lucene Text Indexer; therefore, it reduces the need for memory data structures for caching entity metadata.

#### 3.4 SeRQL Query Engine

The SeRQL Query Engine interface provides the ability to query a triple store repository using SeRQL and to obtain result set through its API. Implementations for sesame serve as wrappers for the vendor-specific APIs, hiding them from the components that use the services.

#### 3.5 Graphical User Interface

The SEMKNOW systems graphical user interface is built on Java Server Pages (JSP). The user interface has been tested and operates optimally with IE9, Mozilla Firefox, Google chrome, Safari 5.

#### 3.6 Semantic Reasoner

A semantic reasoner, reasoning engine, rules engine, or simply a reasoner, is a piece of software able to infer logical consequences from a set of asserted axioms. Reasoner use first-order predicate logic to perform reasoning. The semantic reasoner implemented in system is Pellet.

### 4 IMPLEMENTATION

SEMKNOW, a Semantic web based system is developed with several new practices and technologies as of today. Particular challenges and issues are focused and solved in this system when compared to traditional semantic systems or existing knowledge based semantic systems.

The systems ontology is developed in particular using W3C approved standards through DC (<http://purl.org/dc/elements/1.1>) and BIBO (<http://purl.org/ontology/bibo/>) entries. And Imports of these namespaces makes SEMKNOW ontology standardized.

And the implementation of SEMKNOW system is carried out considering the Indian universities and educational institutions running the digital library repositories on relational databases. The shift from legacy relational databases to semantic knowledge bases with triple storage is made easy through R2O mapping of tables and their columns with entities of ontology [4]. Alignment API is implemented in the system for this purpose by maintaining threshold value of 0.8 for matching from relational database to ontology. Fig 2 shows the R2O mapping between SEMKNOW ontology and relational database.

This approach is of high prior importance in the development phase of the proposed system [5]. It encourages all the developers to move from existing WEB2.0 database driven

<sup>2</sup> Sesame – <http://openrdf.org>

<sup>3</sup> Pellet – <http://clarkparsia.com/pellet>

<sup>4</sup> Protégé – <http://protege.stanford.edu>

applications to Semantic WEB (WEB3.0) database driven applications and build a giant WEB.

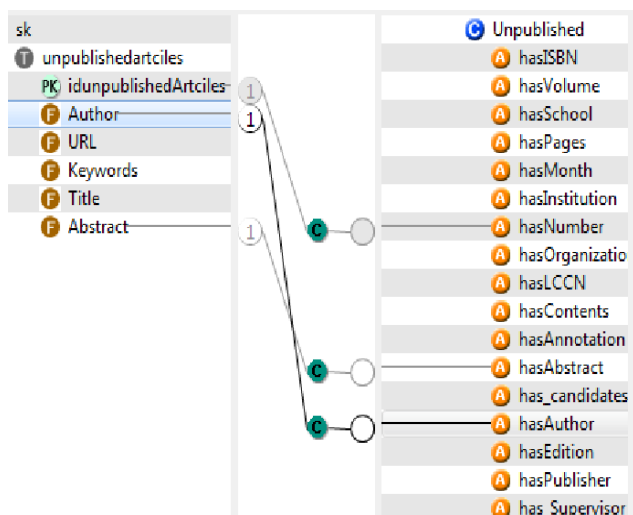


Fig. 2. Mapping between Relational Database and Ontology.

In Fig 2, Left hand-side is the tables and its columns from relational database. And right hand-side is the ontology with its classes and properties.

```
<?xml version="1.0" encoding="UTF-8"?>
<r2o>
  <dbschema-desc name="semdocs">
    <has-table name="journal">
      <keycol-desc name="idJournal"/>
      <nonkeycol-desc name="Year"/>
      <nonkeycol-desc name="Category"/>
      ...
      <nonkeycol-desc name="Keywords"/>
      <nonkeycol-desc name="Abstract"/>
    </has-table>
  </dbschema-desc>
  ...
  </dbschema-desc>
  <conceptmap-def name="file:/C:/Documents%20and%
20Settings/H0208972E/Desktop/sem-original.owl#JournalArticle">
    <uri-as type="DEFAULT">
      <operation oper-id="concat">
        <arg-restriction on-param="string1">
          <has-value>file:/C:/Documents%20and%
20Settings/H0208972E/Desktop/sem-original.owl#JournalArticle</has-value>
        </arg-restriction>
        <arg-restriction on-param="string2">
          <has-column>semdocs.journal.idJournal</has-column>
        </arg-restriction>
      </operation>
    </uri-as>
    <default_uri-as>
      <operation oper-id="concat">
        <arg-restriction on-param="string1">
          <has-value>file:/C:/Documents%20and%
20Settings/H0208972E/Desktop/sem-original.owl#JournalArticle</has-value>
        </arg-restriction>
        <arg-restriction on-param="string2">
          <has-column>semdocs.journal.idJournal</has-column>
        </arg-restriction>
      </operation>
    </default_uri-as>
  </conceptmap-def>
</r2o>
```

Fig. 3. R2O Mapping

The system though being developed through JSP (Java Server Pages) technology for front-end, the standards are validated in security and maintenance through EJB (Enterprise Java Beans) technology.

The challenge in this alignment is API starts troubling as number of records increases. This is dealt with R2O mapping that is efficient even for millions of records.

After successful execution of R2O, the system ontology will be loaded with all records from relational database into ontology as instances.

## 5 SEMKNOW ONTOLOGY

In process of learning by students and researchers various sources of knowledge are being generated in form of publications, articles. Today all this sources of information is available in the internet or in data bases in digital format, which enables their search and use.

But search returns more efficient results only if it is machine understandable. This is achieved through ontologies [6], and so this system is built with specifically designed SEMKNOW application ontology to store and retrieve all the literature documents from various heterogeneous databases.

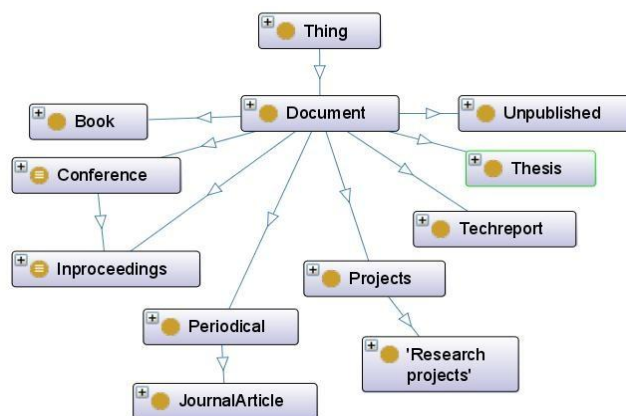


Fig. 4. SEMKNOW Application Ontology.

In Fig 4, the hierarchy is designed through `<rdfs:subClassOf>` and properties (ISBN, Author...) of instances for each Class (Book, Journal Article...) are defined through `<rdfs:datatype>` in RDF.

SEMKNOW is built intelligent enough that when a user deposits a manuscript, the article falls into its category on administrator approval and application ontology is dynamically updated and synchronises with triple store.

It's obvious that there is a tight-bond relation between document and its topic. To facilitate complete effective search the different types of documents organized in application ontology.

gy are related with a suitable domain-or subject-based ontology [7]. So for this purpose SEMKNOW domain ontology has been designed using the ACM digital library hierarchy.

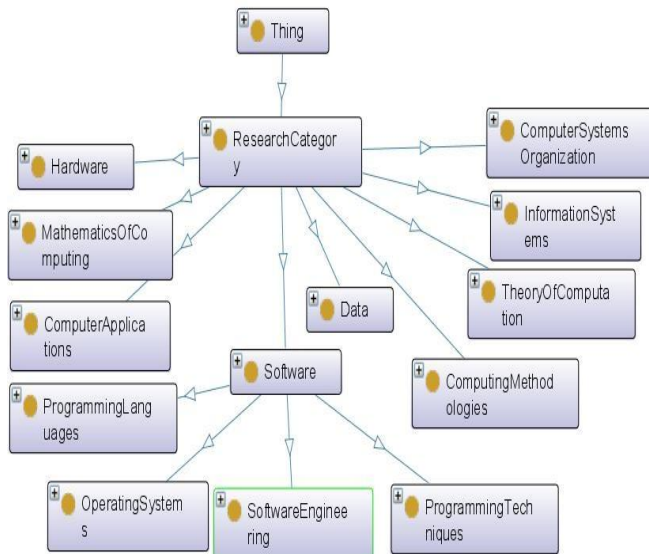


Fig. 5. SEMKNOW Domain Ontology.

Fig 5 represents the domain ontology of the system. The hierarchy in this ontology goes through all the Research categories and narrows down to Subjects in it. The subjects in this ontology are represented as instances through `<rdf:type rdf:resource="http://www.w3.org/2002/07/owl#NamedIndividual">`.

## 6 APPLICATION AND DOMAIN ONTOLOGY INTEROPERATION

The two ontology's are related with `<owl:ObjectProperty rdf:about="&.hasDocs">` in domain ontology and `<owl:ObjectProperty rdf:about="&.isOfSubject">` in application ontology.

These relations are established through Ontology mapping enabling the navigation from TypeOfDocument to SubjectOfDocument and vice-versa.

SEMKNOW System has achieved this with Alignment API. Name eq method from the API solves the integration with 1.0 threshold value [8]. The following is lines from alignment code:

```
xmlns:owl="http://www.w3.org/2002/07/owl#"
<!-- http://sk.owl#ResearchCategory -->
<owl:Class
rdf:about="http://sk_application.owl#ResearchCategory">
<owl:equivalentClass
rdf:resource="http://sk_domain.owl#ResearchCategory">
<rdfs:subClassOf rdf:resource="&owl;Thing"/>
</owl:Class>
```

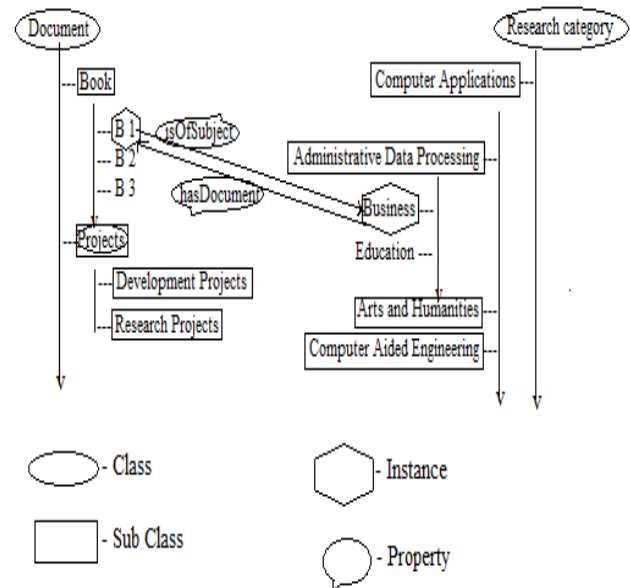


Fig. 6: Concept of SEMKNOWs Application and Domain ontology interoperability.

in Fig 6, the left hand side represents the application ontology and right side is the domain ontology. The instance B1 of class Book has other properties like hasAuthor, hasISBN, hasPublisher., besides shown *isOfSubject* property.

## 7 PROTOTYPE

The SEMKNOW systems prototype developed in Java shows that users perceived the system as useful in assisting research work and the retrieval of information is of adequate quality. The system is built with basic services justifying the objective of the system like browsing and semantic querying [9]. Besides few additional modules are included namely -

### 7.1 Administrator Module

Metadata of new documents can be added and updated to Institutional repository. The system provides user interface for admin to populate triple store without having prior knowledge about adding RDF statements to .owl files every time a new record to be updated, unlike traditional semantic web based systems and handles user activities.

Fig 7 shows the representation of SEMKNOW triple store using Sesame MySQL RDF store [10]. Adding new records in deposit module of system dynamically updates repository



through Sesame Java API.

| Subject        | Predicate          | Object  |
|----------------|--------------------|---|
| SEMKNOW:Author | rdftype            | owl:DatatypeProperty  |
| SEMKNOW:Author | rdfs:subPropertyOf | SEMKNOW:HumanCreator  |
| SEMKNOW:B-1    | SEMKNOW:Author     | "IBM"   |
| SEMKNOW:B-10   | SEMKNOW:Author     | "E.F.Codd"  |
| SEMKNOW:B-11   | SEMKNOW:Author     | "Apple Computer, Inc."  |
| SEMKNOW:B-12   | SEMKNOW:Author     | "Christopher D. Manning, Prabhakar Raghavan and Hinrich Sch  tze" |
| SEMKNOW:B-2    | SEMKNOW:Author     | "John E. Hopcroft; Jeffrey D.Ullman"                              |
| SEMKNOW:B-3    | SEMKNOW:Author     | "O.J.Dahl; E.W.Dijkstra; C.A.R.Hoare"                             |
| SEMKNOW:B-4    | SEMKNOW:Author     | "Alfred V.Aho; Jeffrey D.Ullman"                                  |
| SEMKNOW:B-5    | SEMKNOW:Author     | "Per Brinch Hansen"   |
| SEMKNOW:B-6    | SEMKNOW:Author     | "Per Brinch Hansen"   |
| SEMKNOW:B-7    | SEMKNOW:Author     | "Seymour Papert"  |
| SEMKNOW:B-8    | SEMKNOW:Author     | "Dorothy Elizabeth Robling Denning"                               |
| SEMKNOW:B-9    | SEMKNOW:Author     | "C.A.R.Hoare; C.B.Jones"  |
| SEMKNOW:C-1    | SEMKNOW:Author     | "James Noble; Stuart Marshall; Stephen Marshall; Robert Biddle."  |
| SEMKNOW:C-10   | SEMKNOW:Author     | "Bryan Veal; Annie Foong"   |
| SEMKNOW:C-11   | SEMKNOW:Author     | "Brian A.Barsky"  |
| SEMKNOW:C-12   | SEMKNOW:Author     | "Robert T.Held; Martin S.Banks"                                   |
| SEMKNOW:C-13   | SEMKNOW:Author     | "Koushik Sen"   |
| SEMKNOW:C-14   | SEMKNOW:Author     | "Andrea Arcuri; Xin Yao"  |

Fig. 7. SEMKNOW Repository

## 7.2 Deposit Module

A new document can be added to repository by user but updated to repository only on admins approval.

## 7.3 Discussion Forum

Knowledge Management system is justified on enabling knowledge collaboration and sharing, that is achieved through this module.

Fig 8 represents the SEMKNOW system, displaying records from repository to the end-user.

| Identifier | Title  | Author             |
|------------|--|--------------------|
| C - 7      | Text compression as a test for artificial intelligence         | Matthew V. Mahoney |
| C - 15     | Connect 4 as a problem in artificial intelligence and robotics | Brian L. Stuart    |

Fig. 8. SEMKNOW System.

## 8 CONCLUSIONS AND FUTURE WORK

In this paper, I described the system architecture and general functionality of SEMKNOW, an easy to use semantic web-based application for the scholars, researchers, and students to perform scholarly activities. Current work includes semantic filters enabling content browsing, querying, and knowledge-base development through deposit module in system, system internationalisation with multi-language support depending on geographical location and interactive Knowledge collaboration through discussion forum and blogs.

The ontology engineering greatly helped in enabling the identification of the more significant concepts and relationships used more often in a knowledge domain, thus becoming a prevailing knowledge representation metadata for knowledge repositories. The main contribution of this system has been in developing a Human-Machine System that supports flexible interactions between their components and finally offering the expected information to the end users of SEMKNOW system.

The following stage in the development of this system will be in shifting RDF triple store to Jena TDB that supports uploading, querying millions of triples in petite time compared to current existing triple stores. And also to implement a hybrid reasoner with cached-reasoning service that retrieves pre-computed inferences into the system to overcome expense of time and space when dealing with large ontologies.

## REFERENCES

- [1] Alistair Miles, Brian Matthews, Michael Wilson, Dan Brickley, "SKOS Core: Simple knowledge organisation for the web," Springer Berlin Heidelberg, vol. 5, no. 3, pp. 69-83, DOI: 10.1300/J104v43n03\_04
- [2] Kees van der Sluijs, Geert-Jan Houben, Jeen Broekstra, and Sven Casteleyn. 2006. Hera-S: web design using sesame. In *Proceedings of the 6th international conference on Web engineering (ICWE '06)*. ACM, New York, NY, USA, 337-344. DOI=10.1145/1145581.1145646 <http://doi.acm.org/10.1145/1145581.1145646>
- [3] Jaakko Salonen, Ossi Nyk  nen, Pekka Ranta, Juha Nurmi, Matti Helminen, Markus Rokala, Tuja Palonen, V  nni Alarotu, Kari Koskinen, and Seppo Pohjolainen. 2011. An implementation of a semantic, web-based virtual machine laboratory prototyping environment. In *Proceedings of the 10th international conference on The semantic web - Volume Part II (ISWC'11)*, Lora Aroyo, Chris Welty, Harith Alani, Jamie Taylor, and Abraham Bernstein (Eds.), Vol. Part II. Springer-Verlag, Berlin, Heidelberg, 221-236.
- [4] Jesus Barrasa Rodriguez and Asuncion Gomez-Perez. 2006. Upgrading relational legacy data to the semantic web. In *Proceedings of the 15th international conference on World Wide Web (WWW '06)*. ACM,

New York, NY, USA, 1069-1070. DOI=10.1145/1135777.1136019

- [5] Guntars Bumars and Karlis Cerans. 2010. RDB2OWL: a practical approach for transforming RDB data into RDF/OWL. In *Proceedings of the 6th International Conference on Semantic Systems (I-SEMANTICS '10)*, Adrian Paschke, Nicola Henze, and Tassilo Pellegrini (Eds.). ACM, New York, NY, USA, Article 25, 3 pages.
- [6] Bodo Hüsemann and Gottfried Vossen. 2005. Ontology engineering from a database perspective. In *Proceedings of the 10th Asian Computing Science conference on Advances in computer science: data management on the web (ASIAN'05)*, Stéphane Grumbach, Liying Sui, and Victor Vianu (Eds.). Springer-Verlag, Berlin, Heidelberg, 49-63.
- [7] Janina Fengel and Michael Rebstock. 2009. Model-based domain ontology engineering. In *Proceedings of the 4th International Workshop on Semantic Business Process Management (SBPM '09)*, Martin Hepp, Knut Hinkelmann, and Nenad Stojanovic (Eds.). ACM, New York, NY, USA, 55-58. DOI=10.1145/1944968.1944978
- [8] François Scharffe, Jérôme Euzenat, and Dieter Fensel. 2008. Towards design patterns for ontology alignment. In *Proceedings of the 2008 ACM symposium on Applied computing (SAC '08)*. ACM, New York, NY, USA, 2321-2325. DOI=10.1145/1363686.1364236 <http://doi.acm.org/10.1145/1363686.1364236>
- [9] Daniel Oberle, Steffen Staab, Rudi Studer, and Raphael Volz. 2005. Supporting application development in the semantic web. *ACM Trans. Internet Technol.* 5, 2 (May 2005), 328-358. DOI=10.1145/1064340.1064342
- [10] Ying Yan, Chen Wang, Aoying Zhou, Weining Qian, Li Ma, and Yue Pan. 2008. Efficiently querying rdf data in triple stores. In *Proceedings of the 17th international conference on World Wide Web (WWW '08)*. ACM, New York, NY, USA, 1053-1054. DOI=10.1145/1367497.1367652