Organising web links semantically with Boek

Rakesh Paul & Paul Mc Kevitt
School of Computing & Intelligent Systems, University of Ulster, Magee
John Macrae
Office of Innovation, University of Ulster, Jordanstown
E-mail: paul-r2@email.ulster.ac.uk, p.mckevitt@ulster.ac.uk, j.macrae@ulster.ac.uk

Aims & Objectives

- The aim of this research is to create an open source semantic web framework for developing Rich Internet Applications (RIA).
- Evaluate the framework by developing a semantic web RIA application called Boek (Dutch for "Book") to store favourite web URLs.

Why Semantic Web?

- The Semantic Web (Berners-Lee et al., 2001) and related technologies provides an infrastructure to organise data based on semantics.
- The Semantic Web = a Web with meaning.
- The Semantic Web can describe data in a way understood by machines

Semantic Web Fundamentals

- Resource Description Framework (RDF) (Passin, 2004) - language for describing information and resources making it possible for computer programs to search, collect, analyse and process information.
- Resource Description Framework Schema (RDFS) and Web Ontology Language (OWL) provide vocabularies for expressing metadata.
- RDF describes resources with classes, properties, and values.

Fig. 1. Sample RDF for a web page

- RDF Schema provides the framework to describe application-specific classes and properties.

```
<rdf:RDF>
<rdfs:Class rdf:ID="resources" />
<rdfs:Class rdf:ID="document">
<rdfs:subClassOf rdf:resource="#resources"/>
</rdfs:Class>
</rdf:RDF>
```

Fig. 2. Sample RDFS for an application specific RDF

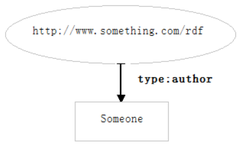

Fig. 3. RDF graph representation of a statement

- Semantic Web provides Logic and Proof (van Harmelen, 2008).
- Logic is used to derive conclusions from a set of relations.
- Proofs trace or explain the steps of logical reasoning.

Flipper is a Dolphin (known relationship)
Every Dolphin is also a Mammal (known relationship)
Flipper is a Mammal (inferred relationship using logic)

Fig. 4. Semantic Web inference from available knowledge

Ontologies

- Ontologies consist of a finite list of terms and relationships between important concepts (classes of objects) of the domain.
- OWL is a superset of RDF and designed to be interpreted by computers.
- SPARQL is a query language and data access protocol for the Semantic Web.
- SPARQL can be used to retrieve information or construct information dynamically from existing RDF graphs.

```
PREFIX boek: <http://www.boek.com/elements/1.0/>
SELECT ?author ?subject ?title ?topicOf
WHERE {
 ?x boek:author ?author ;
 boek:subject ?subject ;
 boek:title ?title ;
 boek:topicOf ?topicOf .
 FILTER ( regex(?title, "jena", "i") || regex(?title, "semantic", "i")
 ) }
```

Fig. 5. Sample SPARQL Query to find documents with details

Semantic Web Applications

- MIT's Piggy Bank (Huynh et al., 2005) facilitates generation and storage of semantic data from existing web pages.
- Piggy Bank extracts information from existing Web pages and stores it in Resource Description Framework (RDF) for publication to a Semantic Bank.
- SemCards (Thórisson et al., 2009) are machine and human-readable entities that enable non-experts to create and use semantic content.

Design & Implementation

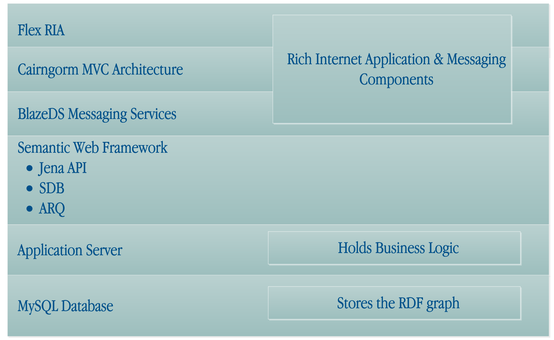

Fig. 6. Boek Architecture

- Software utilised in Boek includes Adobe Flex, Cairngorm, BlazeDS, Jena Semantic Web Framework, SPARQL with Jena SDB and ARQ APIs.
- Jena provides a programming environment for RDF, RDFS , OWL, SPARQL and includes a rule-based inference engine.
- SDB Jena component for storage and querying, specifically for the SPARQL language.
- ARQ is a query engine for Jena that supports the SPARQL RDF Query language. In the Semantic Web querying means reading RDF graphs and is not same as updating the graphs.

```
// Sparql query string
String queryString = "SELECT ?x ?author" WHERE {?x type:author ?author}"

// Creating a Query Object from the query string
Query query = QueryFactory.create(queryString);

// Creating Query execution for
QueryExecution qexec = QueryExecutionFactory.create(query, model);
// model being an RDF graph

// Executing the query against the model
rs = qexec.execSelect() ;

// Iterate the result of query execution
while (rs.hasNext()) {

// Get the query solution for parsing
      QuerySolution qs = rs.nextSolution();

      // parse the query solution for individual result
      parse(qs);
}
```

Fig. 7. Jena ARQ code for querying SPARQL

Boek

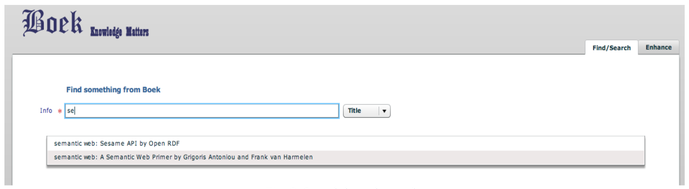

Fig. 8. Search based on title

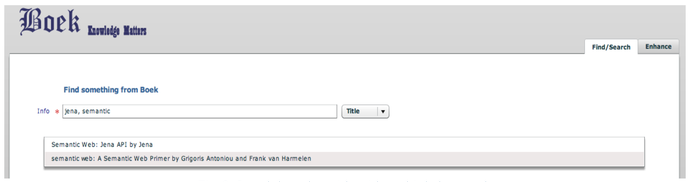

Fig. 9. Search based on title with multiple keywords

Boek

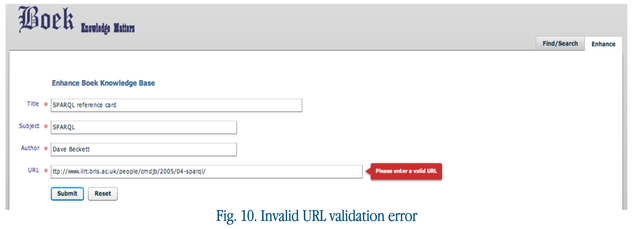

Fig. 10. Invalid URL validation error

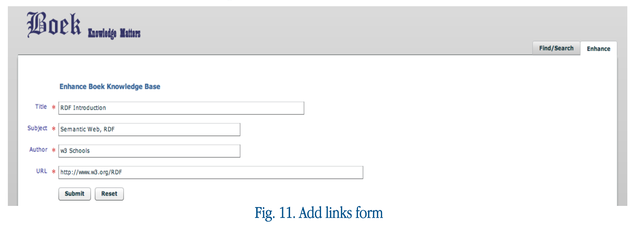

Fig. 11. Add links form

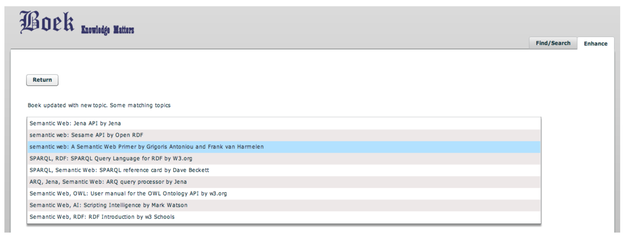

Fig. 12. Adding new Links shows matching topic on subject added

```
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:j.0="http://xmlns.com/foaf/0.1/"
    xmlns:boek="http://www.boek.com/elements/1.0/"
    xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://jena.sourceforge.net/ARQ/">
    <boek:author>Jena</boek:author>
    <boek:subject>ARQ, Jena, Semantic Web</boek:subject>
    <boek:title>ARQ query processor</boek:title>
    <boek:topicOf>http://jena.sourceforge.net/ARQ/</boek:topicOf>
  </rdf:Description>
</rdf:RDF>
```

Fig. 13. RDF data generated while adding new data

Conclusion & Future Work

- Applying the Semantic Web to billions of existing web pages is challenging(Janev & Vranes, 2009).
- Individual web users should be able to create their own semantic web content.
- Boek utilises capabilities of Semantic Web technologies to provide a meaningful bookmarking technique.
- Future work on Boek includes replacing the Flex based GUI with a web browser plug-in.
- Enhance Boek to organise all individual information requirements.
- Use of Ontology learning can extract relevant domain terms from different link' content and personal documents to build personal knowledge repositories. This can enable building clusters of domain data which can be shared between different users.
- Integrate facilities to automatically publish useful links and information from one user to others.

References

- Berners-Lee, T., Hendler, J. and Lassila, O.: The semantic web - a new form of web con-tent that is meaningful to computers will unleash a revolution of new possibilities. In Scientific American 284. (2001) 34–43, May
- Huynh, D., Mazzocchi, S. and Karger, D.: Piggy Bank: Experience the Semantic Web Inside Your Web Browser, Lecture Notes in Computer Science, Vol. 3729. (2005) 413-430
- Janev, V. and Vranes, S.: Semantic Web Technologies: Ready for Adoption?, IT Professional, Vol. 11, No. 5. (2009) 8-16
- Passin, T. B.: Explorer's Guide to the Semantic Web. Greenwich, CT: Manning (2004)
- Thórisson, K.R., Spivack, N. and Wissner, J.M.: SemCards: A New Representation for Realizing the Semantic Web. ICCCI-2009, Wroclaw, Poland, October 5-7. (2009) 425-436
- van Harmelen, F.: Semantic Web Technologies as the Foundation of the Information Infrastructure. In: P. van Oosterom, S. Zlatanove (eds.), Creating Spatial Information Infra-structures: Towards the Spatial Semantic Web, 37-54. London: CRC Press (2008)