THE WOMBRA PROJECT: A WEB-BASED ONTOLOGY-ENHANCED MULTI-PURPOSE BUILDING-REGULATION RETRIEVAL APPLICATION FOR SCOTTISH TECHNICAL STANDARDS

Lewis John McGibbney, BSc (Hons) / PhD Candidate, lewis.mcgibbney@gcu.ac.uk
Bimal Kumar, PhD / Professor of IT in Design and Construction, b.kumar@gcu.ac.uk
School of the Built and Natural Environment, Glasgow Caledonian University, Glasgow, Scotland

ABSTRACT

An increasing amount of professional work within the domain of sustainable design and construction is becoming dependent on retrieving regulatory and advisory information over the web quickly. A common recognition that advanced search and retrieval technologies enable enormous amounts of information to be accessed has emerged. However, as a result designers and builders are finding it increasingly difficult to identify this information and assimilate them in their activities. Commercial search engines provide a generic and inaccurate method of retrieving relevant information for domain-specific needs in a focussed manner. Therefore, there is a need for developing intelligent domain-specific search and retrieval technologies under a structured information management framework. This paper presents Wombra: a web-based information search and retrieval application which employs domain specific ontology to identify (in particular) relevant energy performance building regulations. The framework builds on our previously established model, incorporating an HTTP SPARQL server, which supports the SPARQL RDF query language allowing additional truths to be inferred from our ontologies. Firstly we introduce the design and development of the customised, domain specific web search platform. Related research in this field accompanies this introduction. A theoretical model justifies the choice of technologies used and the basic construction of the search application itself. Further, we provide the reasoning behind our requirement to establish an ontology library with specific application to Scottish Technical Standards. We conclude this paper with various observations relating to on-going experiments and a discussion on future research areas. The Wombra project is being developed in collaboration with a Scottish City Council's building control department who are actively validating the value of our approach in their daily activity of checking and approving designs for construction.

Keywords: Search and Information Retrieval, Ontology, Construction, Building Regulations, Scotland

1 INTRODUCTION

A principal driver for innovation in the area of information systems has been the World Wide Web. Research efforts have focused on the development of systems which make the task of information retrieval and data processing less laborious. There are inherent difficulties when attempting to retrieve data over the web, which consequently provides problems for professionals in general, with the construction industry being no exception. With this in mind, it is unsurprising to note that currently, efficient systems which solve this abstract problem being used in production environments are few and far between. The unprecedented explosion of published data available over the web has resulted in a scenario where users are unable to find, analyse, and utilise information for their own domain specific purpose resulting in a badly structured web of unstructured information. This situation combined with the sheer volume of available data has resulted in the IT community coining the term "information overload". Texts such as (Manning et al. 2008) and (Konchady 2008) provide extensive grounded theory on this high level topic with the latter focusing on specific tools existing under open source licences which can be used to create applications in an attempt to combat the information overload phenomenon. This paper addresses the problem of implicit information contained within web-based building regulations and how we can make this information explicit and therefore directly applicable to the activities of designers and builders.

2 RELATED RESEARCH

Research in this area stems from two sources which share subjective similarities but also maintain disjoint characteristics from one another. To elaborate, it is important to clearly identify and categorise the two strands of literature. The purpose of this exercise is twofold; firstly we remove any analogous perceptions which may create confusion and secondly we merit authors for the significant progress which has been made within discrete research fields. The first strand of research such as (Rezgui et al. 2010) combines the use of semantic technologies and search applications, and the second category relates specifically to the specification, defining syntax and semantics used to express structured queries across data sources as presented in (McCarthy 2005). The former work cited above shares a platform with (Tomassen et al. 2010), as both publications display advanced information retrieval methods through domain specific use of ontology. This specific branch of Tomassen's research focuses on the construction of feature vectors, which are then used within a search application to provide refined search functionality. In both of these cases we see clear evidence of production system implementations with the latter applicable to the Norwegian petroleum industry. Many examples of literature in this field such as (Setchi et al. 2011) diverge to focus on a particular strand of information retrieval, although they converge to draw consensus regarding the potential for improvements in information retrieval results through the use of ontology processing. This opinion is shared by work presented by (Taduri et al. 2009), in which they display the successful implementation of a framework which utilises the use of ontology to improve the retrieval of relevant documents from two independent domains, namely court cases and patents. To conclude this account of related research we focus on formal works such as (McCarthy 2005), and (W3C 2008), definition documents. These contain practical examples of structured queries which aim to infer functional truths from structured RDF (W3C 2004a) data.

3 THE THEORETICAL MODEL

Within this section we focus on three key areas; a summary of the particular problem addressed within the Wombra project, a formal definition of the research aim, and an annotated model of the information framework itself.

The generic problem which we wish to address by developing the Wombra project directly reflects the views expressed in both the introductory material and within the literature review. By means of a series of interviews and a field study questionnaire we were however able to further refine the information overload problem to reflect a particular problem specific to construction and engineering disciplines. In direct contrast to advances in software such as BIM and automated checking of construction design, current software infrastructures which solely focus on information search and retrieval are seriously lacking throughout the construction and engineering sector. We are aware of data resources existing in loosely structured and badly designed clusters or silo's (e.g. independent industry body databases, local authority intranets etc.) over the web. This data is also available in an array of formats, consequently making the job of retrieving specific data for construction design activities a far from trivial task. Further to this, considering the scenario that we were able to retrieve the correct document(s), specifically locating sections/clauses from extensively used unstructured proprietary data formats such as MS Word, PDF, etc. has still not been successfully implemented within current search architectures. The problem of making implicit data contained within proprietary data formats explicit and therefore easily retrievable by domain users has still to be addressed. The prototype system we have in development provides a means to address all of the above issues through the use of ontology processing for enhanced search and retrieval functionality placing emphasis on the retrieval of implicit domain specific knowledge through use of structured queries which are processed by an HTTP SPARQL server named Fuseki (Jena 2011).

3.1 Research definition

Definitively an accurate description of the Wombra project would be as follows

Definition 1 *"To develop a web-based information search and retrieval application which employs domain specific ontology to identify and infer truths from (in particular) relevant Scottish energy performance building regulations."*

The nature of the project has permitted the addition, testing and analysis of various components within the framework. However at its core, the fundamental application functionality remains within the category of information search and retrieval. At an abstract level the process of modelling is to provide a formal representation of a designed or actual object, process or system, or a representation of a reality. Our early work aimed at assembling a theoretical framework enabled a thorough analysis of essential features and their specific functionality in respect to the aim of the research project.

Figure 4 included in Appendix 1 displays the key components of the Wombra framework. In essence, it provides two main strands of functionality; query refinement provided by an ontology enhanced search mechanism which comes in the form of an JavaScript widget, and secondly a SPARQL (W3C 2008) query platform enabling users to infer additional information from RDF graphs existing in persistent storage. Nutch (ASF 2011), a top level project licensed by the Apache Software Foundation, is an open source web-search project written in Java which builds on existing search architectures adding web-specifics, such as a crawler, a link-graph database and parsers for HTML and other document formats. We have automated the crawler to undertake an internet crawl which once completed delegate's data storage to a Solr index. Integration of the ontology enhanced query refinement comes in the form of a JavaScript widget built using Jena; a Java based programming toolkit for building semantic web applications. The widget implements an RDF/OWL parser to parse any ontology models provided (in this case ontology files residing within our ontology library) and retrieves all subclasses and instances of entities within the ontology. These are then enhanced with synonyms from the Word Net (Princeton 2011), corpus, to find possible entities with semantic equivalence within crawled web pages. When a user submits a query the widget retrieves information of relevance from the ontology stored within the web-based ontology repository, matches documents within the Solr index and presents document matches accompanied by a list of similar documents comprising of synonymic counterparts. Progressing to the second strand of functionality provided by the framework: Fuseki, an HTTP server which supports the SPARQL query language for RDF. This enables users to submit structured queries and retrieve specific RDF results, in their preferred syntax, based upon an array of criteria. This additional functionality allows users to infer additional truths about concepts within our ontologies. An example scenario would be a query submitted asking for the value of a specific building element's data type property such as the U-value of roof light glazing within a domestic dwelling. This request would be converted to a SPARQL query which would then be processed by Fuseki and we would retrieve the data values from knowledge encoded within our ontologies. Some early work undertaken, showing examples of SPARQL queries are included in our experiments and results in Section 5.

4 ONTOLOGY LIBRARY

In a similar approach to the one adopted in (Rezgui et al. 2010), the process of gathering end user requirements prior to construction of an information portal, accompanied by the continuous consultation process throughout software design and development was the methodology used in this research.

4.1 A methodology for construction and engineering ontologies

By placing emphasis upon social and environmental problems (Rezgui 2010) which need to be considered prior to the development of any information processing software application we were able to ensure integrity within the design of Wombra. Factors included: the heterogeneous and highly fragmented nature of construction industry, dynamic formation of project teams and business relationships whom may never work together again, labour intensive nature of industry and growing requirement for skilled workforce as IT becomes more sophisticated, and the highly regulated nature of industry with numerous governing bodies existing over a widespread geographic area, to name a few. All of the aforementioned present significant and valid arguments which are fundamental topics for consideration prior to the development of both a tailored solution through use of software and secondly the establishment of any domain specific ontology library.

It is important to consider that any given ontology resource should maintain the following attributes; firstly concepts encoded within the ontology should be constructed by both experts in the field and potential end-users, past experience has shown a clear requirement for widespread use of some set of standard vocabulary within construction and engineering, therefore any ontology should recognise and utilise domain specific vocabulary throughout its application. Within Scottish technical standards this is exactly the case as regulations are created by a panel comprising of certifiers and verifiers. Decisions are then made based on the principle that individuals are suitably qualified and sufficiently experienced within the field and can therefore be responsible for decision making. Finally, ontologies should be developed with extensibility at the forefront of their design. As industry changes, ontological resources should shadow these changes. Scottish building regulations are dynamic documents and are subject to periodic change dependant on a number of variables. These characteristics are directly applicable to Scottish energy performance regulations in particular as Scottish ministers have adopted a range of procedures which directly affect regulatory material namely; a staged approach towards reaching net zero carbon emissions for space heating, hot water, lighting and ventilation within the next ten years, if practical, the ambition of total-life zero carbon buildings by 2030 and finally the consideration of zero fees for building warrant applications where new buildings are to be significantly above the current energy standards. The most recent regulations were published in October 2010 and in keeping with the points raised above which are included in the independent (Sullivan Report 2007), the Scottish Government has formally committed to further review energy standards within both domestic and non-domestic dwellings for 2013 and 2016 respectively. Figure 1 shows a snippet from one of our ontologies displaying the named individual *#exampleGlazingToConservatory*. From the RDF/XML we can see that the individual represents a previously constructed class *#Glazing*. This individual also maintains two rdf:datatype values representing direct values encoded within the Scottish technical standards 2010 for domestic dwellings, section 6 energy (Scottish Government 2010). The rdf:datatypes are of primitive data type *#double*, both of which individually represent the values of different glazing configurations for domestic dwellings.

```
- <!--
    http://www.semanticweb.org/ontologies/2010/11/section6energydomestic2010.owl#exampleGlazingToConservatory
  -->
- <owl:NamedIndividual rdf:about="http://www.semanticweb.org/ontologies/2010/11/section6energydomestic2010.owl#exampleGlazingToConservatory">
    <rdf:type rdf:resource="http://www.semanticweb.org/ontologies/2010/11/section6energydomestic2010.owl#Glazing"/>
    <hasMaximumAreaWeightedAverageUvalueForAllElementsOfTheSameTypeWperm:K rdf:datatype="http://www.w3.org
   /2001/XMLSchema#double">2.0</hasMaximumAreaWeightedAverageUvalueForAllElementsOfTheSameTypeWperm:K>
    <hasMaximumIndividualElementUvalueWperm:K rdf:datatype="http://www.w3.org
   /2001/XMLSchema#double">3.3</hasMaximumIndividualElementUvalueWperm:K>
  </owl:NamedIndividual>
```

Figure 1: Code snippet of named individual *#exampleGlazingToConservatory*

The choice of methodology behind ontology design is very much dependant on the nature and characteristics of the targeted domain and its various applications, as well as the resources and development time available and the required depth of analysis of the ontology (Rezgui 2007). The degree of detail included within our ontologies directly reflects the specific problems highlighted throughout earlier sections of this paper and the justification contained within Definition 1.

4.2 The anatomy of a SPARQL query

Before progressing to early experiments and results, it is important to understand why we may wish to query ontologies within our ontology library. Although we utilise the OWL 2 Web Ontology Language (W3C 2009), as a representational format for displaying meaning within domain concepts, in order to infer additional information, or truths, within Fuseki, we are required to use (amongst others) the RDF/XML (Resource Description Framework/eXtensible Mark-up Language), (W3C 2004b), syntactic specification format. OWL data in this format provides a standardised method of representing information over the web; this makes it suitable and convenient for use within semantic web applications. The inherent problem with inferring RDF data is that it can be difficult to decipher the information we require, from the syntactic jungle of RDF we may be presented with. To aid our understanding of data by means of decreasing its visual complexity we utilise a “triple-like” serialization referred to as Turtle (W3C 2011), the Terse RDF Triple Language. This format simulates the underlying characteristics associated with linked data e.g. subject, predicate, object. This is to say that every topical subject relates to some other entity object by means of a linking mechanism (predicate). If for example we were to consider a simple **door** (subject) which in its current state is open, say we wished to **shut** (being the object) the door, the linking mechanism would be the process of closing (to **close**) which we can therefore refer to as the predicate. By using this model we begin to create relationships or links between concepts creating truth in the process. Further to this, Turtle provides a distinct extension of N-Triples (W3C 2004c); used to express RDF test cases and to define the correspondence between RDF/XML and the RDF abstract syntax, whilst carefully taking the most useful and appropriate features added from Notation 3 (Berners-Lee 2009), such as compaction and readability. In essence, this retains the RDF model whilst providing important attributes for end-users. The ultimate purpose of the above is the consequent ability to utilise the Turtle language within an RDF query engine, allowing us to form '*queries by example*', using the data to make an initial query which can then be edited to use variables where bindings are required (amongst others). Hence making full use of the expressiveness of the OWL 2 constructs for formalising properties and data values associated with concepts.

Generally speaking, the recurring use of domain specific grammar combined with the structurally complex nature of building regulations (as well as numerous other legislative documents) means that they are both technically difficult to understand and inherently challenging to navigate. Consequently, problems associated with the above characteristics include document users dedicating long times actually locating specific regulations required to satisfy specific design issues, inaccurate and inconsistent use of domain specific terminology; which in itself has knock-on effects such as misunderstanding between communication, and failure to actually extract full meaning from the regulatory information source; this leaves it to users to infer truths from legislative regulations based on their individual knowledge competency within the specific field, to name but a few. The movement to provide a mechanism which would provide solutions to the above problems in a user oriented, structured, reliable and accurate manner provides justification behind the incorporation of an RDF HTTP SPARQL query engine into the Wombra project. The ability to submit SPARQL queries enables us to formally assert truths about concepts contained within our ontologies.

5 EXPERIMENTS AND RESULTS

This section discusses our early experimental results focused on asserting truths from RDF datasets stored within our ontology library. Within this section we focus on two different types of structured SPARQL queries, both of which present interesting information retrieval results for discussion. This prepares a platform for further consideration in the concluding section of this paper. As we focus solely on SPARQL queries within this paper the reader is directed to (McGibbney et al. 2011), for previous experiments regarding text based ontology refined queries also possible under the Wombra framework. In order to undertake these experiments it was necessary to utilise the Fuseki, SPARQL server, which provides a stable infrastructure for obtaining results using the SPARQL protocol over HTTP. The technically challenging nature of building regulations mean that domain specific knowledge is usually encoded in a latent manner. Users without industry knowledge and the experience of navigating such documents find that inferring this information is a far from trivial task. Our test sample; Section 6 – Energy from the Scottish technical standards 2010 provided us with information which we have coded within our ontology in a number of ways.

5.1 A simple SPARQL query

The example shown in Figure 1 displays numeric data values associated with concepts, however as described above there are several other types of data associated with concepts, such as a narrative commentary existing in string text form and domain ranges to name a few. In plain English, the query shown in Figure 2 seeks to find all entities within an rdfs:range which have a linking predicate variable and are bound to an rdfs:comment (which in this case is of data type string). The results from this simple query can be seen in Table 1.

```
PREFIX xsd:     <http://www.w3.org/2001/XMLSchema#>
PREFIX rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:    <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl:     <http://www.w3.org/2002/07/owl#>
PREFIX fn:      <http://www.w3.org/2005/xpath-functions#>
PREFIX apf:     <http://jena.hpl.hp.com/ARQ/property#>
PREFIX dc:      <http://purl.org/dc/elements/1.1/>
PREFIX ns:      <http://www.semanticweb.org/ontologies/2010/
11/section6energydomestic2010.owl#hasNoMaximumOrMinimumAreaGuidance>

SELECT ?range ?b ?comment
WHERE
{  ?b rdfs:range ?range; rdfs:comment ?comment. }
```

Figure 2: Simple SPARQL query

In practice we actually retrieve several other subclasses within the range for this particular query along with their associated object property and text string data value. In plain English, the data results in Figure 3 display that the building element *Roof Window* has no maximum or minimum area guidance. This information is then accompanied by the narrative included in the column titled 'comment'. A query of this nature highlights at an abstract level clear evidence that we are accurately able to retrieve implicit data from our RDF dataset.

Table 1: Snippet of SPARQLer query results for simple query

SPARQLer Query Results		
range	b	comment
<http://www.semanticweb.org/ontologies/2010/11/section6energydomestic2010.owl#RoofWindow>	<http://www.semanticweb.org/ontologies/2010/11/section6energydomestic2010.owl#hasNoMaximumOrMinimumAreaGuidance>	"Due to the target method set by carbon dioxide emissions standard 6.1, there is no need for guidance on minimum or maximum area for windows, doors, rooflights and roof windows in new dwellings."^^<http://www.w3.org/2001/XMLSchema#string>

5.2 Inferring numeric RDF data resources

Our second query has been structured specifically for the purpose of retrieving numeric data values from individuals existing within our RDF datasets. Figure 3 displays the query designed to retrieve data from the RDF triple excerpt shown in Figure 1. In this example we utilise the FILTER operand to request data which has a numeric data value of less than 2.5. The intent is to display the powerful benefits of using structured queries on structured datasets; we are easily able to obtain more precise results based upon our required level of granularity. In English the query requests three variables; any named individual of a specific type owl:NamedIndividual bound to an object by a predicate where the object is filtered to have a numeric data value of less than 2.5. Results can be seen in Table 2. In this case the query results display that the NamedIndividual *example Glazing to Conservatory* has a maximum area weighted average U-value for all elements of the same type with a numeric data value of 2.0 Wperm^{2}K.

In agreement with the first of our queries, the data inferred from the triples in Figure 1 also displays accurate retrieval results. From our on-going experiments we have learned that the use of various constructors within SPARQL queries provide a wealth of options to any application user depending directly upon their data retrieval requirements. The example queries chosen for inclusion in this paper constitute a small but significant portion of the type of data in regulatory information we wish to make explicit using the Wombra framework. The queries display results for the retrieval of both narrative as well as numeric data encoded within Scottish building regulations. Much of this information is usually encoded in a latent fashion and therefore either easily unacknowledged or else misinterpreted, which in turn results in incorrect decisions being made when checking designs for regulatory compliance. Although we demonstrate that it is possible to change the nature of this data making it explicit, accessible and therefore easily applicable to design checking processes the trivial examples above raise additional questions which merit further discussion, some of which are addressed in section 6.

```
PREFIX xsd:     <http://www.w3.org/2001/XMLSchema#>
PREFIX rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:    <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl:     <http://www.w3.org/2002/07/owl#>
PREFIX fn:      <http://www.w3.org/2005/xpath-functions#>
PREFIX apf:     <http://jena.hpl.hp.com/ARQ/property#>
PREFIX dc:      <http://purl.org/dc/elements/1.1/>
PREFIX ns:      <http://www.semanticweb.org/ontologies/2010/
11/section6energydomestic2010.owl#exampleGlazingToConservatory/>
SELECT ?NamedIndividual ?p ?v WHERE
{
    ?NamedIndividual rdf:type owl:NamedIndividual.
    ?NamedIndividual ?p ?v.
    FILTER (?v < 2.5)
}
```

Figure 3: Numeric SPARQL query

Table 2: Snippet of SPARQLer query results for numeric query

SPARQLer Query Results		
NamedIndividual	p	v
<http://www.semanticweb.org/ontologies/2010/11/section6energydomestic2010.owl#exampleGlazingToConservatory>	<http://www.semanticweb.org/ontologies/2010/11/section6energydomestic2010.owl#hasMaximumAreaWeightedAverageUvalueForAllElementsOfTheSameTypeWperm^{2}K>	"2.0"^^<http://www.w3.org/2001/XMLSchema#double>

6 DISCUSSIONS & FURTHER RESEARCH

This research paper documents a portion of our research agenda included under the framework we refer to as the Wombra project. We have highlighted our early efforts aimed at establishing a web-based information search and retrieval application with specific application to Scottish building regulations, specifically Section 6 – Energy, Scottish Technical Standards 2010. This section serves its purpose for further discussion in an attempt to identify limitations and initial conclusions we can draw from the research framework to date. The aforementioned problems we seek to address have prompted investigation into the engineering discipline, construction process, quality requirements and ultimately the final product with regards to both development of our ontology library and our software infrastructure. Fundamentally, within the Wombra project we propose a solution aimed at solving the high level problem which exists under the umbrella term 'information overload'. Additionally our search for more robust and accurate information retrieval methods which address issues of a construction domain specific nature has resulted in the construction and on-going validation of a structured information management procedure which acts in parallel with a full text-based ontology enhanced search platform. To our knowledge, within Scotland a framework of this nature is unique in both its design and functionality; the remainder of this section expands to consider some of the criteria we have identified as essential requirements for individual verification of our approach.

We recognise that a query language such as SPARQL is technically challenging both to learn and to implement in an accurate and time effective manner. The high complexity of structured queries mean that even if all relevant knowledge could be transformed into structured form, many users would still have great difficulties when attempting to retrieve this information (Magatti et al. 2010), therefore we are working to accommodate the complete SPARQL infrastructure within our existing servlet container, which could then be accessed via an unobtrusive interface. The behavioural characteristics associated with typical user activity indicate that non-standard searches on information sources e.g. use of a structured search technique, is generally not welcomed by typical users. Typically, SPARQL queries require deep formal thinking and a sound knowledge of the way in which data is stored within the data store. Our intention is to establish a SPARQL query store to accompany the search architecture. This would provide a tailored out-of-the-box solution which would work in parallel with the interface to enable users who are unfamiliar with structured search techniques an effective platform to execute powerful queries linked to building regulations. During the course of this work, we consulted both data curators and end users e.g. the Scottish Government's building standards division and members of South Ayrshire Council's building control department. Interestingly this resulted in us observing a stark contrast between the way in which curators want their data to be used and think that it is being used, and the actual methods being adopted within building control departments. This observation concluded that although these parties and the activities which they carry out are independent, the consequences and resulting outcomes of the way in which building regulations e.g. Scottish technical standards are produced means that the processes are inextricably linked. This lays a platform for further analysis and research into the inherent nature of processes at both ends of the spectrum. Our early hypothesis states that a better understanding of this concept is required to ensure integrity and consistency within the planning and design checking processes which are directly affected by the information encoded within building regulations. Further validation of Wombra is currently underway by firstly comparing the text-based search with the structured SPARQL information retrieval method, before finally progressing to benchmark our system against other commonly used methods for the retrieval of energy performance building regulations which are then used during design checking and regulatory compliance processes.

ACKNOWLEDGMENTS

The authors would like to thank members of the South Ayrshire Council's building control department as well as members of the Scottish Government's building standards division who are assisting us in validating and improving our framework.

APPENDIX A

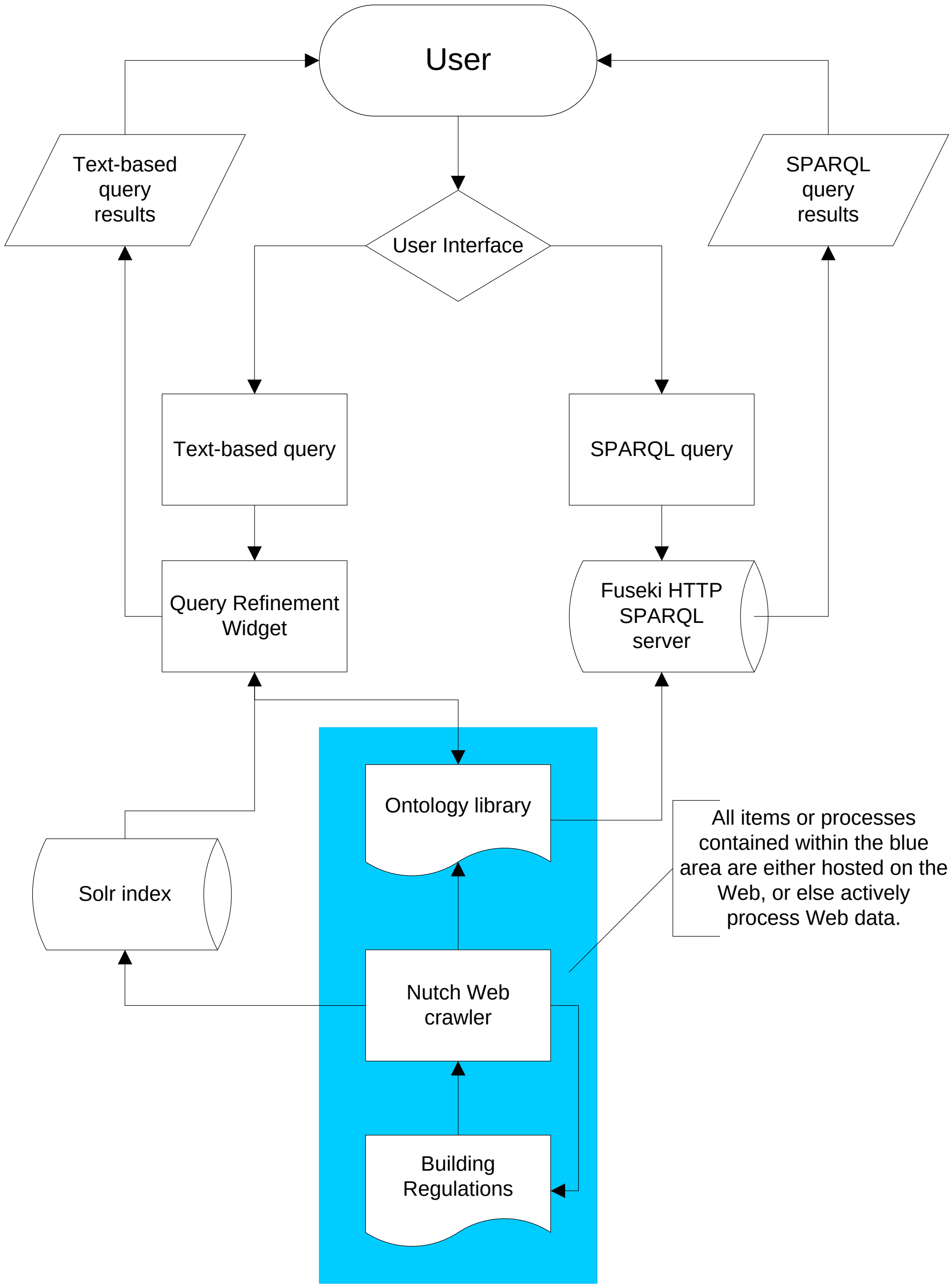

Figure 4: The Wombra research framework

REFERENCES

Apache Software Foundation. (2010) "Welcome to Nutch.", last accessed 02/02/2011 at http://nutch.apache.org

Berners-Lee, T. (2009) "Notation 3 An readable language for data on the Web.", last accessed 25/04/2011 at http://www.w3.org/DesignIssues/Notation3

Jena. (2011) "Fuseki – SPARQL Server." last accessed 04/05/2011 at http://openjena.org/wiki/Fuseki

Konchady, M. (2008) "Building Search Applications: Lucene, Lingpipe, and Gate.", Mustru Publishing, Oakton, Virginia.

Magatti, D., Steinke, F., Markus, B., Tresp, V. (2010) "Combined Structured and Keyword-Based Search in Textually Enriched Entity-Relationship Graphs" *First Workshop on Automated Knowledge Base Construction*, Paper #2, Grenoble, France.

Manning, C. D., Raghavan, P., Schütze, H. (2008) "An Introduction to Information Retrieval.", Cambridge University Press, Cambridge, England.

McCarthy, P. (2005) "Search RDF data with SPARQL – SPARQL and the Jena Toolkit open up the semantic web." *An article from IBM developerWorks*, last accessed 19/04/2011 at http://www.ibm.com/developerworks/xml/library/j-sparql/

McGibbney, L. J., Kumar, B., (2011) "A Knowledge-directed Information Retrieval and Management Framework for Energy Performance Building Regulations." *2011 ASCE International Workshop on Computing in Civil Engineering*, June 19-22, Miami, FL, USA.

Princeton. (2010) "The WordNet homepage." last accessed 12/26/2010 at http://wordnet.princeton.edu/

Rezgui, Y., Wilson, I. E., Miles, J., Hopfe, C. J. (2010) "Federating information portals through an ontology-centred approach: A feasibility study." *Advanced Engineering Informatics*, 24 (3). 340-354.

Rezgui, Y. (2007) "Text-based domain ontology building using Tf-Idf and metric clusters techniques." *Knowledge Engineering Review*, 22 (4). 379-403.

Setchi, R., Tang, Q., Stankov, I. (2011) "Semantic-based information retrieval in support of concept design." *Advanced Engineering Informatics*, 25 (2). 131-146.

Sullivan, L. (2007) "A Low Carbon Building Standards Strategy for Scotland." *Report of a panel appointed by Scottish Ministers Chaired by Lynne Sullivan*, arcamedia publishing.

Taduri, S., Hang, Y., Lau, G., Law, K., Kesan, J. P. (2009) "Developing a Comprehensive Patent Related Information Retreival Tool" *Journal of Theoretical and Applied Electronic Commerce Research*, 4 (1). 1-15.

The Scottish Government. (2010) "Technical Handbooks 2010 Domestic Section 6 Energy" last accessed 18/04/2011 at http://www.scotland.gov.uk/Resource/Doc/217736/0096961.pdf

Tomassen, S. L., Strasunskas, D. (2010) "Measuring intrinsic quality of semantic search based on feature vectors." *International Journal of Metadata, Semantics and Ontologies*, 5 (2). 120-133.

W3C (2009) "OWL 2 Web Ontology Language Document Overview." *W3C Recommendation 27 October 2009*, last accessed 29/04/2011 at http://www.w3.org/TR/owl2-overview/

W3C (2004a) "Resource Description Framework (RDF)."last accessed 22/04/2011 at http://www.w3.org/RDF/

W3C (2004b) "RDF/XML Syntax Specification (Revised)." *W3C Recommendation10 February 2004*, last accessed 26/03/2011 at http://www.w3.org/TR/rdf-syntax-grammar/

W3C (2004c) "RDF Test Cases." *W3C Recommendation 10 February 2004*, last accessed 17/04/2011 at http://www.w3.org/TR/rdf-testcases/#ntriples

W3C (2008) "SPARQL Query Language for RDF" *W3C Recommendation 15 January 2008*, last accessed 29/04/2011 at http://www.w3.org/TR/rdf-sparql-query/

W3C (2011) "Turtle – Terse RDF Triple Language." *W3C Team Submission 28 March 2011*, last accessed 02/05/2011 at http://www.w3.org/TeamSubmission/turtle/