Chapter 7
Semantically Enhanced Search and Browse

Alistair Duke and Jörg Heizmann

Abstract  Squirrel, a search and browse tool that provides access to semantically annotated data is described. The tool offers a hybrid approach to search allowing the user to enter simple search terms and then refine their search or browse to related material through the presentation of appropriate metadata. Squirrel builds upon and integrates a number of the semantic technology components described elsewhere in the book. These include machine learning and information extraction components which generate, extract and manage semantic metadata contained within and about textual documents at index time. A number of run-time components have also been integrated to deliver an enhanced user experience such as natural language generation which provides natural language summaries of knowledge held in formal (ontological) structures; device independence which allows the tool to be run on multiple devices; result consolidation which presents the most relevant textual content of result documents rather than a simple list of results; and a natural language interface which translates natural language queries into structured queries formulated with respect to a given ontology.

7.1 Introduction

This chapter describes Squirrel, a search and browse tool that provides access to semantically annotated data. Search is seen as a key application that can benefit from semantic technology with improvements to recall and precision versus conventional Information Retrieval techniques. Squirrel builds upon and integrates a number of the semantic technology components described elsewhere in this book. These include machine learning and information extraction components which generate, extract and manage semantic metadata contained within and about textual documents at index time. A number of run-time components have also been
integrated to deliver an enhanced user experience such as natural language generation which provides natural language summaries of knowledge held in formal (ontological) structures; device independence which allows the tool to be run on multiple devices; result consolidation which presents the most relevant textual content of result documents rather than a simple list of results; and a natural language interface which translates natural language queries into structured queries formulated with respect to a given ontology.

A hybrid approach has been adopted to provide a balance between the speed and ease of use of simple free text search and the power of semantic search. Users are able to enter simple search terms and then refine their search or browse to related material through the presentation of appropriate metadata. Alternatively, users can express a query using natural language which is analysed and converted into a semantic query.

The chapter is structured as follows. The next section introduces a scenario which illustrates both the requirements and benefits of enhanced search and browse. Section 7.3 describes the salient features of the supporting components that comprise Squirrel and presents an architecture for the system. In Sect. 7.4 further detail is provided about the user experience with a description of the important features of the interface. Section 7.5 discusses related work and outstanding issues whilst we conclude in Sect. 7.6 with a view of the way forward.

### 7.2 Scenario

Squirrel has been trialled in a case study which is developing an improved Digital Library for BT, as discussed elsewhere in this volume. The following scenario describes a Digital Library user carrying out a knowledge seeking task, making use of the features provided by Squirrel and its supporting components.

The user has an initial high level goal to seek information about the field of Home Health Care and has little idea about what is contained in the library which might be of use. He first enters “Home Health Care” into the search box and hits the “Go!” button. The first result screen includes a summary of the sorts of resources that have been found that might be able to meet his needs. These include the textual resources from the library that is that there are a number of journal articles, conference papers, periodicals and web pages (shared by library users and indexed as library resources) that match his query. In addition, there a number of library topics that also match his query including one which is itself called “Home Health Care”. Further matches include a number of organisations from a domain knowledge base whose description includes the search term.

In addition to this summary, the user is also presented with the top ranked textual resources (in typical search engine style). This simple list of documents is augmented with the ability to refine the search based on the properties of the documents in the result set, including a hierarchical display of the topics of the results documents, date of publish, author name, etc.