Introducing the Gnowsis Semantic Desktop

Leo Sauermann, Sven Schwarz
DFKI GmbH Knowledge Management Group
Kaiserslautern, Germany
{ leo.sauermann | sven.schwarz }@dfki.de

Abstract

We present our vision of a Semantic Web enhanced desktop environment, the *Semantic Desktop*. Common desktop resources like MS-Outlook contacts or Mozilla bookmarks are treated as Semantic Web Resources. The *Gnowsis* system is our open source project that implements this vision to be used and extended by developers.

1 Introduction

The Semantic Web is still in its beginning. We have to push forward and start building useful applications. The Semantic Desktop approach does not focus on organizational or worldwide systems, it's focus is to organize data on personal computers. The use of Semantic Web technology in desktop applications will create a benefit for users. When users benefit from it, the technology will be used and the Semantic Web will grow. Tim Berners-Lee tells us from time to time to start writing applications [Berners-Lee 2003].

On a modern personal computer there are thousands of documents, emails, photos and music songs stored. Their number increases as users download new files, send and receive emails, socialize and exchange contact information. A normal application can handle only part of the whole data.

All these information items can be treated as Semantic Web resources. Existing technology and approaches from the Semantic Web are then used on these desktop resources. Metadata of as well as relationships between all the resources on a computer are represented in the Resource Description Framework (RDF). Data from different applications are linked using triples. A Semantic Browser is used to visualize the information and allow users to benefit. This results in a Semantic Web enhanced Desktop – the *Semantic Desktop*. Applications still handle only their part of the whole – and access the other parts through RDF.

There are different approaches in this topic. Haystack [Quan *et al.*, 2003] is similar to Gnowsis. Haystack is also a framework to build RDF applications, having its own user interface and RDF engine. In the Gnowsis project, we decided to build an architecture to integrate existing applications instead of replacing them. In today's software projects it is a common demand that new software integrates with Microsoft's Outlook or popular tools like the Mozilla suite. We created a framework to integrate different applications. Using this data integration framework, it is possible to extract information on the fly from common applications. Developers can also extend the framework and integrate custom data sources.

The Semantic Desktop will play a key role in building a global Semantic Web. First users will annotate their personal data, motivated by the benefits of an integrated personal information management. Then this annotated data can be published on the web.

Web sites don't have to be annotated with the "add semantic sugar on top" philosophy, instead all resources can be annotated starting with the time of creation.

Once a file is annotated, it will be a small step to add the annotations when the file is sent through email, posted on a website or stored in an organizational memory.

1.1 Everything is a Resource

To build such a system, an experimental approach in the Semantic Web field had to be taken. While people still discuss the meaning and interpretation of URIs on the RDF-IG newsgroup and the W3 [Berners-Lee 2002], we tested different approaches and implemented them. All resources on the desktop are identified by a URL, be it a file, an email or a database row. These URLs can be used to annotate the resources using third party applications. The framework parses the URLs to find the resources.

There are many ways to exchange RDF data between applications or how to query them from servers. Available implementations are the Joseki Server[1] by HP Labs or Sesame[2] which both provide HTTP interfaces. We decided to use RDQL[3] to find resources and an

[1] http://www.joseki.org/
[2] http://www.openrdf.org/
[3] http://www.hpl.hp.com/semweb/rdql.htm

extension of Concise Bounded Descriptions (CBD) [Stickler 2004] to retrieve the RDF data.

1.2 Implementation

The popular RDF framework Jena[4] by HP Labs was chosen as the main technical ground to build Gnowsis on. We extended the Jena framework with the Semantic Desktop functions. For different data sources like the file-system or Microsoft Outlook we created adapters, which are accessible as Jena models. We also integrated existing adapters, for relational databases we used *D2RQ* from Chris Bizer [Bizer 2004]. Developers and Researchers can use these Jena Models in their own applications to extract information from the different data sources.
Developers can access the functionality through XML/RPC and RMI interfaces. The framework is Java based and adapted to Windows. We are working on Linux integration.

2 System Design

The architecture will be explained using a typical use case. The system is designed for typical tasks of today's users, using common applications and the popular MS-Windows operating system.

2.1 Typical Use Case

A typical use case is that the user wants to know more information about a music file. Right-clicking the file allows the user to "browse" in his personal Semantic Web. The request is sent to the local Gnowsis server passing the URL of the file and the "browse" command identifier. It parses the URL and identifies it as a file.
There are several adapters registered in the system, one handles local files and can extract file-size and name. Another adapter can extract the ID3-Tag information of the MP3 file, like artist and track number. Each adapter is shipped with an RDFS vocabulary describing the possible values that the adapter can extract. The central server queries both adapters about the MP3 file. Each adapter extracts his part of the information. Also, a central RDF database is queried to identify related resources or additional annotations (like "I received this file from Peter – the foaf:Person from my address book").
The result of this query is integrated and again accessible through a Jena model. Finally, the server starts a user interface module, the semantic browser, to visualize the result.
In the browser, the user can see information about the resource and manipulate the resource. Clicking on the resource tells the server to open the file using the standard application for music. A second resource can be selected in another application (like a person in Outlook) and linked to the resource.

[4] http://jena.sourceforge.net/

2.2 RDF Integration

We provided several alternatives to implement adapters. The first way is to implement a "find(subject, predicate, object) method (that allows wildcards). For this, we created a framework that helps writing the code.
The second way is through *"Chatty Bounded Descriptions"*, a variation of Patrick Stickler's CBD [Stickler 2004]. To get information about a resource, only the URL of it is passed to the adapter. As a result, the adapter returns a subgraph around the resource. The result is passed as RDF/XML string or Jena Model. Time consuming parts are first left out and can be returned on explicit request.

About the project

The Gnowsis system was first created in Vienna by Leo Sauermann as his diploma thesis. Through discussions with developers from the RDF-IG and researchers from the German DFKI, new features and use cases where developed. At the moment several research projects of the DFKI are enriched using Gnowsis.
The project is hosted and documented on the website http://www.gnowsis.com. In September 2004 we release an alpha version. The source code and Java binaries will be available under an open source license. Supported platforms are MS-Windows (were all features work), Linux, and MacOsX.
We will continue the development and plan to create better user interfaces. We also plan to implement collaboration features in order to create a Networked Semantic Desktop.

References

[Berners-Lee 2003] Tim Berners-Lee, *Where to Direct Our Energy* ,2nd International Semantic Web Conference, October 23, 2003
http://www.w3.org/2003/Talks/1023-iswc-tbl/Overview.html

[Berners-Lee 2002] Tim Berners-Lee, "What do HTTP URIs identify?", July 2003
http://www.w3.org/DesignIssues/HTTP-URI

[Bizer 2004] Chris Bizer, *D2RQ - Treating Non-RDF Databases as Virtual RDF Graphs*
http://www.wiwiss.fu-berlin.de/suhl/bizer/d2rq/index.htm

[Quan *et al.*, 2003] Dennis Quan, David Huynh, and David R. Karger. *Haystack: A Platform for Authoring End User Semantic Web Applications*. International Semantic Web Conference 2003: 738-753
For more references, see the URLs in the footnotes.

[Stickler 2004] Patrick Stickler. *URIQA the Uri Query Agent Model*. Contains also the methodology of Concise Bounded Descriptions (CBD)
http://sw.nokia.com/uriqa/URIQA.html