The Information Workbench
Interacting with the Web of Data

Peter Haase[1], Andreas Eberhart[1], Sebastian Godelet[1], Tobias Mathäß[1],
Thanh Tran[2], Günter Ladwig[2], Andreas Wagner[2]

[1]fluid Operations GmbH, Walldorf, Germany
{firstname.lastname}@fluidops.com
[2]Institute AIFB, University of Karlsruhe, Germany
{gla,dtr,awa}@aifb.uni-karlsruhe.de

Abstract. We present the Information Workbench, an application for interacting with the Web of data. The Information Workbench manages large amounts of structured and unstructured information, which may be imported and integrated from existing sources, but also allows end users to annotate, complete and update information in a collaborative way. New paradigms for accessing information include hybrid search across the structured and unstructured data, keyword search combined with facetted search, as well as semantic query completion and interpretation, which assists the user in expressing complex information needs by an automated translation of keyword queries into hybrid queries. A *Living UI* based on widgets for the interaction with the data enables a homogeneous, seamless, continuous and personal experience.

1 Main Features

In recent years, we have observed a tremendous success of the paradigms of the Web 2.0 in Web applications. The Web has developed from a platform in which information is published by few providers to an interactive and collaborative medium for producing, consuming and sharing information. As the most prominent example, Wikipedia has grown to one of the central knowledge sources.

At the same time, the amount of structured data available on the (Semantic) Web has been increasing rapidly. Currently, there are billions of triples published and connected in web data sources of different domains. The benefits of this linked data are obvious (at least to our community), but there are still very few applications that actually make use of it. In particular, the potential of applications complementing the Web of data with the characteristics of the Web 2.0 has been largely unrealized: Existing applications are mainly limited to generic linked data browsers, they typically assume that the data is published by data providers – and thus read only – as apposed to user generated content. In other words, the means for the interaction with the Web of data are still in its 1.0 state.

The Information Workbench provides the means to fill this gap – as an infrastructure for building applications for the interaction with the Web of data, combining Web 2.0 features such as collaboration with those of semantic technologies. The key features of the Information Workbench include:

– the ability to manage large amounts of structured and unstructured content, which may be imported and integrated from existing sources, but also may be generated by end users, who can annotate, complete and update the content,
– new paradigms for accessing information, including hybrid search across the structured and unstructured data, keyword search combined with facetted search, and semantic query completion and interpretation, automatically translating keywords into hybrid queries corresponding to the user intent,
– a *Living UI* to enable a homogeneous, seamless and personal experience, despite heterogeneous and dynamic data. Knowing what, when, and where to show is realized through an automated and customizable selection of widgets, which implement various paradigms for interacting with the data.

The technology of the Information Workbench is generic in the sense that it is independent of particular domain or data set, or application. In fact, the strength lies in the ease of building concrete applications. To demonstrate this, we have setup an instance of the Information Workbench to interact with a Semantic Wikipedia, publicly accessible at `http://iwb.fluidops.com/`. To bootstrap the system, we have taken the English Wikipedia and enriched it with structured data from the Open Linked Data Initiative, including the DBpedia data set.

While the data in the demonstrator spans many domains (in fact it covers a large fraction of the world knowledge) and thus potential applications are just as manifold, we further illustrate the benefits in a small application scenario.

2 An Application Scenario

Sebastian is a hobby astronomer, who is familiar with using computers, experienced in using Web 2.0 style wikis and forum software for managing pictures and observation reports. Especially in the domain of astronomy, information is abundant. For example, Wikipedia contains vast amounts of knowledge about astronomy. Most of this knowledge is available in unstructured form, but increasingly, structured data is published. As one example, DBpedia already contains some structured knowledge about astronomic entities extracted from Wikipedia. Using the Information Workbench, Sebastian is able to access and interact with the data aggregated from the various available sources. The data is presented in a resource-centric way, with a single page per resource. One such page may be that of the solar system (c.f. Figure 1). For displaying the Information, the application automatically selects appropriate widgets based on the data available. For example, cosmic objects might be associated with coordinates. Based on them, these objects are displayed using Google Sky. At the same time, Sebastian would like to personalize the interface to his preferences: Sebastian may want to have a Twitter feed included that displays live news about a particular resource, while some other user may prefer to see videos associated with that resource.

An important means of interacting for Sebastian is the ability to enrich the existing data, in the form of photographs, text, and also, more structured data. In a simple case, he may want to annotate the unstructured information within the Wikipedia. Simple annotations such as making the relationship between two entities explicit (e.g. stating `[[location::Solar System]]` for a planet)

Fig. 1. Screenshot of the Demonstrator

lead to immediate benefits: Promptly, Sebastian will be able to perform an ad-hoc structured query, e.g. asking for the mass and planets in the solar system arranged in the order of the distance from the sun. The Information Workbench will assist in formulating complex information needs by automatically translating the query from keywords into structured queries and making suggestions for completing and refining the query. Results are displayed using an appropriate visualization, e.g. in the form a bar diagram type.

Figure 1 shows a screenshot the Information Workbench in our scenario. The screen shows the page of the solar system, integrating information from the original Wikipedia page, the DBpedia data set, various external sources and the annotations added by Sebastian. The upper left widget shows a wiki-based interface to the resource, the widget below shows the associated structured as a graph. On the right side, we see external content in the form of video from YouTube and a Twitter news feed. The upper right widget shows the results of a structured query associated with the page, displaying the density of the planets in the solar system.

For more details on the search concepts implemented by the Information Workbench, we refer the reader to [1].

References

1. Thanh Tran, Peter Haase, and Rudi Studer. Semantic search - using graph-structured semantic models for supporting the search process. In *ICCS*, pages 48–65, 2009.