Semantic Content Processing in Web Portals

Felicitas Löffler*, Bahar Sateli†, Birgitta König-Ries*, René Witte†
*Institute for Computer Science, Friedrich-Schiller University of Jena, Germany
†Department of Computer Science and Software Engineering, Concordia University, Montréal, Canada

***Abstract*—Web portals provide a standardized way of integrating multiple information sources and applications in a single web interface. However, they currently do not provide semantic support for users that need to navigate the often overwhelming amount of content. We demonstrate our open source portal architecture "hanüwa" that integrates text mining web services, based on the Semantic Assistants framework, with the Liferay portal server.**

I. Introduction

Web portals are a specific kind of web-based systems that provide for an integration of diverse information sources and applications. Deployed for a concrete scenario in an organization, they typically address the information needs of a wide range of users and their tasks through both internal and external services.

While a web portal provides convenient access to information, there is no standardized way that allows to further process the available content in order to support users in their tasks. There is also a lack of appropriate technologies for document filtering within a web portal. We envision a new generation of web portals that can provide context-sensitive support through semantic analysis services, in particular based on natural language processing (NLP). These services are deployed in shared or private servers and can be dynamically requested by users that ask for help in a specific task: e.g., finding entities in a documents, summarizing a text, answering a question, or linking content to external sources. As such, they perform the role of AI "assistants" that support their users. Furthermore, we imagine enhancing web portals with a personalization component to adapt the content to the user's needs. Sorting documents or highlighting terms according to a specific user interests would be a great advantage for the user and a step towards working against information overload.

In previous work, Bakalov et al. [1] demonstrated the feasibility and usability of a portal integration with natural language processing services. However, this implementation was tied to a specific, commercial portal engine (IBM WebSphere[1]). The work presented here is a complete re-design and re-implementation of the NLP-portal integration, taking into account future extensions and based exclusively on open source software. Similar to the solution presented in [1], we rely on the Semantic Assistants framework [2] for brokering text mining pipelines as web services, but our new architecture is based on the Liferay[2] open source portal server.

Our new *portlets* can be deployed in any existing Liferay-based portal to offer natural language processing services to its users. Here, we demonstrate the core functionality with named entity recognition in a given article, but the framework is not limited to a single domain: A clear separation of concerns allows a language engineer to make new NLP services available without requiring knowledge in portal technology, and a web engineer can easily design a new web portal that incorporates language technology.

II. Architecture

Our novel Semantic Assistants-portal integration architecture, illustrated in Fig. 1, is designed to allow various portlets to benefit from NLP techniques on their content. The core idea is to enable generic portlets to communicate with the *Semantic Assistants portlet*, specifically designed to connect to the back-end Semantic Assistants server and provide inquiry and invoking capability of NLP pipelines to portal users.

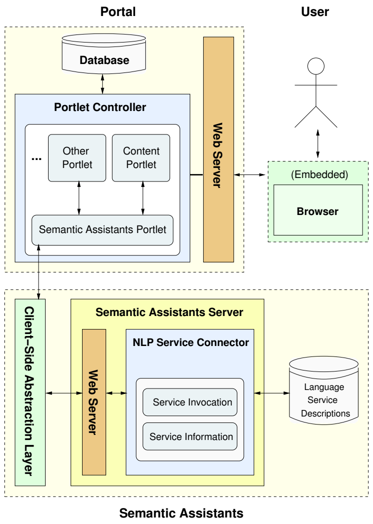


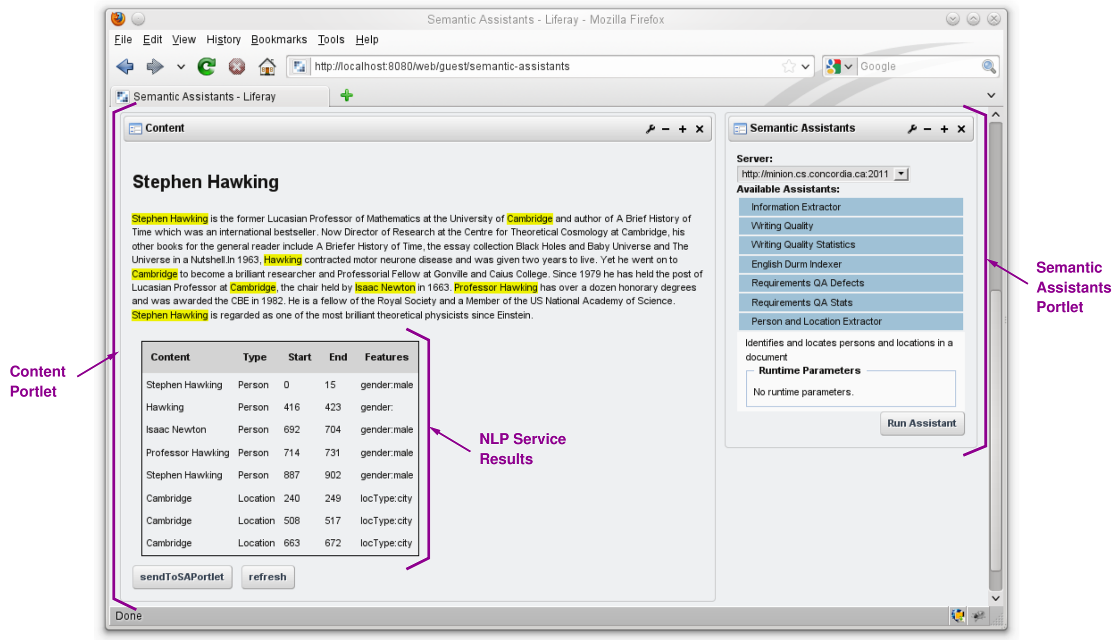
Fig. 1. The Semantic Assistants-Portal Integration architecture

In this architecture, all available portlets in a page can communicate with the Semantic Assistants portlet by sending content for analysis and receiving the results. To commence an analysis session, users interact with the portal via their web browser, for example, on their desktop computer or from a mobile device. Through this integration, users can select an NLP service to execute on a portlet's content from a dynamically-generated list of available *assistants* in the Semantic Assistants server repository. Where applicable, users can also customize the services' behaviour by setting runtime parameters. An execution request is then sent to the Semantic Assistants server from the Semantic Assistants portlet in form of a W3C[3] standard web service call that triggers the execution of the designated NLP pipeline on the provided content. The results of each

[1] IBM WebSphere, http://www.ibm.com/software/websphere

[2] Liferay, http://www.liferay.com/

[3] World Wide Web Consortium (W3C), http://www.w3.org


Semantic Assistants - Liferay - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://localhost:8080/web/guest/semantic-assistants
Google
Semantic Assistants - Liferay
Content
Semantic Assistants
sendToSAPortlet
refresh
Run Assistant
Done
Content Portlet
NLP Service Results
Semantic Assistants Portlet

Stephen Hawking

Stephen Hawking is the former Lucasian Professor of Mathematics at the University of Cambridge and author of A Brief History of Time which was an international bestseller. Now Director of Research at the Centre for Theoretical Cosmology at Cambridge, his other books for the general reader include A Briefer History of Time, the essay collection Black Holes and Baby Universe and The Universe in a Nutshell.In 1963, Hawking contracted motor neurone disease and was given two years to live. Yet he went on to Cambridge to become a brilliant researcher and Professorial Fellow at Gonville and Caius College. Since 1979 he has held the post of Lucasian Professor at Cambridge, the chair held by Isaac Newton in 1663. Professor Hawking has over a dozen honorary degrees and was awarded the CBE in 1982. He is a fellow of the Royal Society and a Member of the US National Academy of Science. Stephen Hawking is regarded as one of the most brilliant theoretical physicists since Einstein.

Content	Type	Start	End	Features
Stephen Hawking	Person	0	15	gender:male
Hawking	Person	416	423	gender:
Isaac Newton	Person	692	704	gender:male
Professor Hawking	Person	714	731	gender:male
Stephen Hawking	Person	887	902	gender:male
Cambridge	Location	240	249	locType:city
Cambridge	Location	508	517	locType:city
Cambridge	Location	663	672	locType:city

Server: http://minion.cs.concordia.ca:2011

Available Assistants:

- Information Extractor
- Writing Quality
- Writing Quality Statistics
- English Durm Indexer
- Requirements QA Defects
- Requirements QA Stats
- Person and Location Extractor

Identifies and locates persons and locations in a document

Runtime Parameters

No runtime parameters.

Fig. 2. Semantic Assistants-Portal Integration User Interface in Liferay

successful service execution are first received by the Semantic Assistants portlet and then passed on to the portlet that requested the service execution. The NLP pipelines are described in the OWL[4] language and the Semantic Assistants server uses SPARQL[5] for a dynamic discovery of available services upon each user request. Hence, adding or removing NLP services to the integration requires no modification to the code base of the portal.

The basis of the personalization component will be an ontology-based user profile, where all user interests are recorded automatically while browsing through the portal and reading documents. A user interface, embedded into a portlet, allows a user to control interests, add new terms, delete or change concepts. The user can also enable or disable the personalization mode. When personalization is desired, the documents are re-sorted and the relevant terms of the user profile are highlighted within the text. In contrast to [1], the personalization feature will be available to various portlets in form of services, rather than a concrete implementation on a per-portlet basis.

III. Application

The integration of NLP assistants within a portal context allows for a multitude of applications. Fig. 2 shows an example scenario in which a portal user needs assistance in analyzing the textual content available in the content portlet (left). Such assistance can be offered to the user through the NLP services listed in the Semantic Assistants portlet (right). This portlet allows the user to connect to different Semantic Assistants servers and review the list of their available pipelines in order to find a suitable assistant for his task at hand. In our example, the list of assistants contains a “Person and Location Extractor” service that extracts entities of *person* and *location* types from a given text. The user then sends the text in the content portlet to the Semantic Assistants portlet for analysis and requests the service execution by clicking on the “Run Assistant” button. This interaction will request the designated Semantic Assistants server for the execution of the ANNIE pipeline, provided by GATE.[6] Subsequently, the results are returned to the content portlet in form of *annotations* in a tabular format and highlighted in the text based on their offsets. The processing time for different scenarios depends on both the length of the input text and the actual NLP pipeline. Naturally, sophisticated NLP pipelines with deep syntactic or semantic analysis require more time to process. Currently, we are working on a personalization scenario aimed at tackling the user’s information overload issue, by filtering the portal’s content according to a user’s interest. The idea is to embed such capability directly within portlets, allowing users to be able to switch to various personalization modes.

IV. Conclusions

In this paper, we described our open source integration of natural language processing capabilities within a portal environment. We also intend to integrate a personalization feature into portals to adapt their content according to a user’s needs. Furthermore, we want to provide a user interface to give the users the opportunity to have control over their recorded interests. The NLP-portal integration will be available as part of the Semantic Assistants distribution hosted on SourceForge.[7]

References

[1] F. Bakalov, B. Sateli, R. Witte, M.-J. Meurs, and B. König-Ries, “Natural Language Processing for Semantic Assistance in Web Portals,” in *IEEE International Conference on Semantic Computing (ICSC 2012)*. Palermo, Italy: IEEE, September 2012.

[2] R. Witte and T. Gitzinger, “Semantic Assistants – User-Centric Natural Language Processing Services for Desktop Clients,” in *3rd Asian Semantic Web Conference (ASWC 2008)*, ser. LNCS, vol. 5367. Bangkok, Thailand: Springer, Feb. 2–5, 2009 2008, pp. 360–374. [Online]. Available: http://rene-witte.net/semantic-assistants-aswc08

[4] Web Ontology Language, http://www.w3.org/2004/OWL/

[5] SPARQL Query Language, http://www.w3.org/TR/rdf-sparql-query/

[6] General Architecture for Text Engineering (GATE), http://gate.ac.uk/

[7] Semantic Assistants, http://sourceforge.net/projects/semantic-assist/