Semantic Assistants Wiki-NLP Showcase

- Semantic Assistants
- Semantic Wiki
- NLP

Introduction

As part of the Semantic Assistants project, we developed the idea of a "self-aware" wiki system that can develop and organize its own content using state-of-art techniques from the Natural Language Processing (NLP) and Semantic Computing domains. This is achieved with our open source Wiki-NLP integration, a Semantic Assistants add-on that allows to incorporate NLP services into the MediaWiki environment, thereby enabling wiki users to benefit from modern text mining techniques.

Here, we want to exhibit how a seamless integration of NLP techniques into wiki systems helps to increase their acceptability and usability as a powerful, yet easy-to-use collaborative platform. We hope this will help you to identify new human-computer interaction patterns for other scenarios, allowing you to make the best possible use of this new technology. A short summary is also available in our WikiSym demo paper [1], Sateli, B., and R. Witte, "Supporting Wiki Users with Natural Language Processing", *The 8th International Symposium on Wikis and Open Collaboration (WikiSym 2012)*, Linz, Austria : ACM, 08/2012. You can also look at our SMWCon 2013 presentation slides [2], Sateli, B., and R. Witte, "Smarter Wikis through Integrated Natural Language Processing Assistants", *Semantic MediaWiki Conference (SMWCon) Spring 2013*, New York City, NY, USA, 03/2013.

Background

Natural Language Processing (NLP) is a branch of computer science that employs various Artificial Intelligence (AI) techniques to process content written in natural language. NLP-enhanced wikis can support users in finding, developing and organizing knowledge contained inside the wiki repository. We realized this idea by developing a comprehensive architecture that offers novel NLP solutions within a wiki environment through a user-friendly interface.

The Wiki-NLP integration user interface embedded in a wiki page

With this interface, users can easily execute sophisticated NLP analysis pipelines on wiki content. The NLP analysis results can be formatted and stored in the wiki in different, user-definable locations (original page, discussion page, new page, etc.)

NLP-enhanced Wikis: Case Studies

We present three different deployments of our Wiki-NLP integration and show how various NLP services can help users in their tasks: (1) Cultural Heritage Data Management (DurmWiki), (2) Biomedical Knowledge Curation (*Intelli*GenWiki), and (3) Software Requirements Engineering (ReqWiki).

Cultural Heritage Data Management in DurmWiki

Cultural heritage data of a society, such as books, are often preserved in a digitized format and stored in distributed repositories. Such a body of content can be turned into a knowledge base accessible to both humans and machines using modern techniques from the Semantic Computing domain. Our DurmWiki contains a digitized version of a German historical encyclopedia of architecture. As browsing and keyword-based search are the only information retrieval means of a classical wiki system, discovering significant knowledge is a major challenge for users of the heritage data. This is further compounded by the fact that these texts contain outdated terminology no longer in use, thus cannot be found through keyword search.

Within DurmWiki, we integrated an NLP service that can perform automatic indexing of a wiki's content, storing it in the wiki itself, similar to classical back-of-the-book indexes. The generated index page, as shown in the figure below, presents an alphabetically-ordered list of terms found in the wiki, together with a direct link to the pages inside the wiki where they appear. In

experiments with end users, we found that the presence of such an automatically maintained index page not only aggregates the wiki's embodied content on a high level and enables users to find information at-a-glance, but also helps them to "discover" interesting concepts or entities that they did not know were present in the wiki.

Automatic back-of-the-book index generation for wiki content

For more information, please refer to the Durm project page and [3], [4]. You can also browse a (limited) demo version online, where you can see the automatically generated index.

Biomedical Knowledge Curation in *Intelli*GenWiki

*Intelli*GenWiki is our solution for life science researchers that need to analyze and manage large amounts of research publications, e.g., for knowledge curation tasks.

By storing research publications and other relevant textual content in a wiki, entities of interest can be automatically detected and added to each page through our Wiki-NLP integration. For example, using our open source OrganismTagger, each species name can be automatically detected and added to the corresponding wiki page, together with additional information, such as the scientific name or a link to the NCBI Taxonomy Database.

Content	Type	Start	End	Features
Magnaporthe oryzae	Organism	143	161	■ NCBI_Taxonomy_WebPage: http://www.ncbi.nlm.nih.gov/Taxonomy /Browser/wwwtax.cgi?id=318829&mode=info ■ organism_scientific_name: Magnaporthe oryzae ■ organism_alias: Magnaporthe oryzae ■ google_search: http://www.google.com /search?q=Magnaporthe+oryzae ■ NCBI_Taxonomy_ID: 318829

Entity detection example in *Intelli*GenWiki

In addition to automatic information extraction from wiki content, *Intelli*GenWiki implicitly produces semantic metadata that can be exploited in various ways, e.g., to be exported to external repositories or to provide semantic entity retrieval capabilities in the wiki, where applicable. In the figure below, we illustrate how curators using *Intelli*GenWiki can find wiki pages containing specific entities of their interest based on their type.

Property:Enzyme

⋈ Page Name	⋈ Enzyme Entities Found
PMID: 20709852	Cellobiohydrolase Cellulases endoglucanases β-glucosidases Invitrogen DNA polymerase

Semantic Entity Retrieval in *Intelli*GenWiki
In a user study with expert biologists, the NLP support in the wiki reduced the curation time by nearly 20%. For more information, please refer to our *Intelli*GenWiki page, as well as [5] and [6].

Collaborative Software Requirements Engineering with ReqWiki

Software requirements engineering is the process of eliciting and documenting the needs of various stakeholders of a software project. Wikis, as an affordable, lightweight documentation and distributed collaboration platform, have demonstrated their capabilities in requirements engineering processes.

However, because of the lenient structure of wikis and the natural language that is used in software requirements specifications (SRS), the presence of semantic defects, such as ambiguity or vagueness, in SRS documents is inevitable. ReqWiki is a novel open source web-based approach based on a semantic wiki that includes our NLP assistants, which work collaboratively with humans on the requirements specification documents. In ReqWiki, users can invoke various generic or domain-specific quality assurance NLP services on the SRS documents using the Wiki-NLP user interface, in order to detect and amend the extracted defects. The figure below shows the results of a readability and a writing quality analysis service invoked on a use case document excerpt.

Pre-Conditions	The manager must be identified and authenticated in the application
Success end condition	The tasks is created and assigned to the technicians with status Assigned.

Readability Metrics on UC/Manage_Tasks (View)

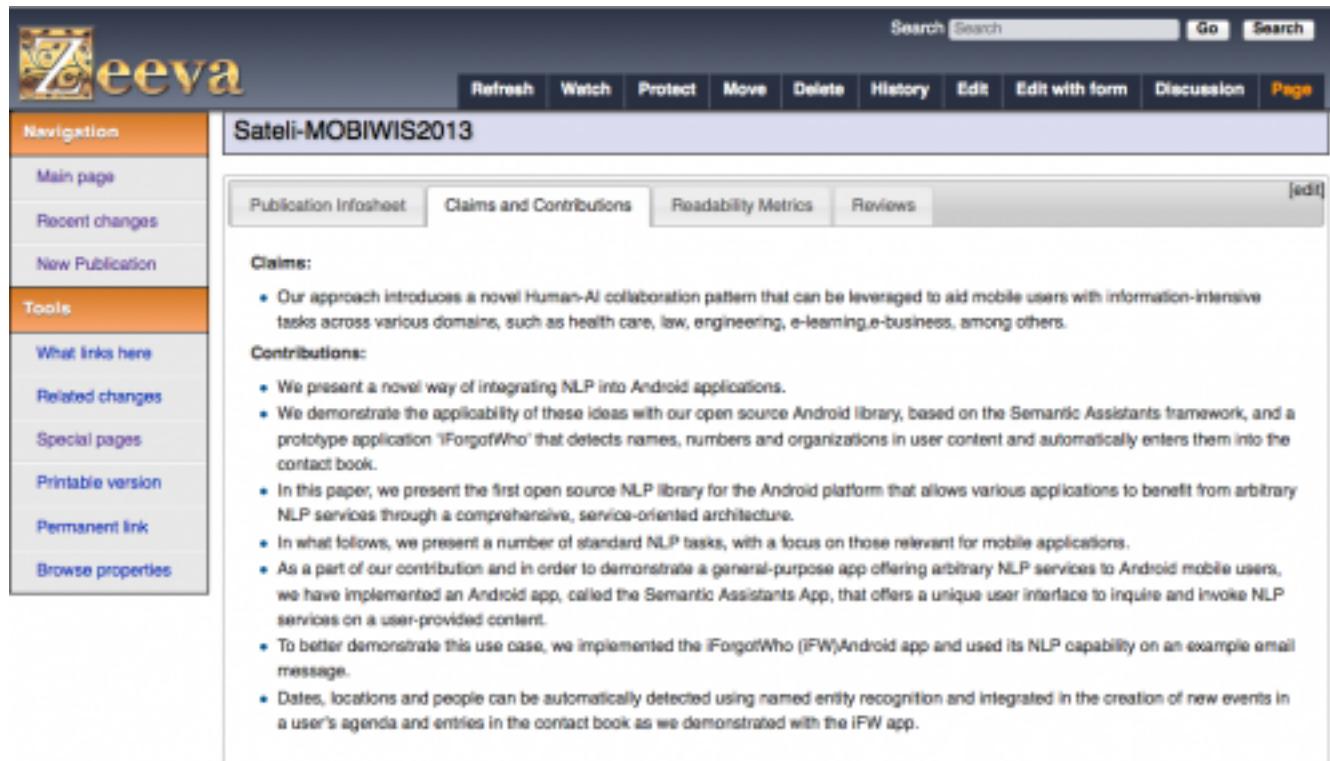| Content	Type	Start	End	Features
The tasks is created and assigned to the technicians with status Assigned.	Passive Voice	686	760	▪ The sentence has been detected as passive and can be improved by changing the verb phrase

Writing Quality on UC/Manage_Tasks (View)

Content	Type	Start	End	Features
The tasks is	Grammar	686	698	▪ problem: Wrong Auxiliary Verb ▪ suggestion: The task is

Automatic Quality Assurance of Wiki Content in ReqWiki

Within a user study with a large number of software engineering students, we confirmed that our assistants can be easily used without prior knowledge of NLP, and that they can significantly improve the quality of a specification, compared to a version without NLP support [7]. For more information, please refer to our ReqWiki page and our publications [7], [8], and [9].

Semantic Research Literature Management with Zeeva

This overabundance of literature available in online repositories is an ongoing challenge for scientists that have to efficiently manage and analyze content for their information needs. Most of the existing literature management systems merely provide support for storing bibliographical metadata, tagging, and simple annotation capabilities. Here, we go beyond these approaches by demonstrating how an innovative combination of semantic web technologies with natural language processing can mitigate the information overload by helping in curating and organizing scientific literature. *Zeeva* is our latest research prototype for demonstrating how we can turn existing papers into a queryable knowledge base.

Semantic infosheet of bibliographical metadata extracted from a sample paper

In Zeeva, users can upload the scientific publications to the wiki, where they are analyzed by a number of NLP pipelines. One of these detects rhetoric entities like *claims* and *contributions* in the publications, which is then transformed into semantic markup that can be queried from other pages. This allows Zeeva to create summary "factsheets" for each paper, which show the detected claims and contributions, as well as other NLP generated content, like an index or a writing quality assessment.

For more information, please refer to our Zeeva page, as well as our publications [10], [11], [12], [13].

Further Information

Please refer to the following pages for more information:

- The main page describing the Wiki-NLP integration
- An overview for the Semantic Assistants
- We also have a tutorial on NLP in wikis that was last presented at SMWCon Spring 2014 [14].

References

1. Sateli, B., and R. Witte, "Supporting Wiki Users with Natural Language Processing", *The 8th International Symposium on Wikis and Open Collaboration (WikiSym 2012)*, Linz, Austria : ACM, 08/2012.
2. Sateli, B., and R. Witte, "Smarter Wikis through Integrated Natural Language Processing Assistants", *Semantic MediaWiki Conference (SMWCon) Spring 2013*, New York City, NY, USA, 03/2013.
3. Witte, R., R. Krestel, T. Kappler, and P. C. Lockemann, "Converting a Historical Architecture Encyclopedia into a Semantic Knowledge Base", *IEEE Intelligent Systems*, vol. 25, no. 1, Los Alamitos, CA, USA : IEEE Computer Society, pp. 58--66, January/February, 2010.
4. Sporleder, C., A. van den Bosch, and K. Zervanou (Eds.), Witte, R., T. Kappler, R. Krestel, and P. C. Lockemann, "Integrating Wiki Systems, Natural Language Processing, and Semantic Technologies for Cultural Heritage Data Management", *Language Technology for Cultural Heritage* Springer Berlin Heidelberg, pp. 213--230, 2011.

5. Sateli, B., C. Murphy, R. Witte, M. - J. Meurs, and A. Tsang, "Text Mining Assistants in Wikis for Biocuration", *5th International Biocuration Conference*, Washington DC, USA : International Society for Biocuration, pp. 126, 04/2012.

6. Sateli, B., M. - J. Meurs, G. Butler, J. Powlowski, A. Tsang, and R. Witte, "IntelliGenWiki: An Intelligent Semantic Wiki for Life Sciences", *NETTAB 2012*, vol. 18 (Supplement B), Como, Italy : EMBnet.journal, pp. 50–52, 11/2012.

7. Sateli, B., E. Angius, S. S. Rajivelu, and R. Witte, "Can Text Mining Assistants Help to Improve Requirements Specifications?", *Mining Unstructured Data (MUD 2012)*, Kingston, Ontario, Canada, October 17, 2012.

8. Sateli, B., S. S. Rajivelu, E. Angius, and R. Witte, "ReqWiki: A Semantic System for Collaborative Software Requirements Engineering", *The 8th International Symposium on Wikis and Open Collaboration (WikiSym 2012)*, Linz, Austria : ACM, 08/2012.

9. Sateli, B., E. Angius, and R. Witte, "The ReqWiki Approach for Collaborative Software Requirements Engineering with Integrated Text Analysis Support", *The 37th Annual International Computer Software & Applications Conference (COMPSAC 2013)*, Kyoto, Japan : IEEE, pp. 405–414, 07/2013.

10. Sateli, B., and R. Witte, "Collaborative Semantic Management and Automated Analysis of Scientific Literature", *The 11th Extended Semantic Web Conference (ESWC 2014)*, vol. 8798, Anissaras, Crete, Greece : Springer, pp. 494-498, 05/2014.

11. Sateli, B., and R. Witte, "Semantic MediaWiki (SMW) for Scientific Literature Management", *The 9th Semantic MediaWiki Conference (SMWCon Spring 2014)*, Montréal, Canada, 05/2014.

12. Castro, A. G., C. Lange, P. W. Lord, and R. Stevens (Eds.), Sateli, B., and R. Witte, "Supporting Researchers with a Semantic Literature Management Wiki", *The 4th Workshop on Semantic Publishing (SePublica 2014)*, vol. 1155, Anissaras, Crete, Greece : CEUR-WS.org, 05/2014.

13. Sokolova, M., and P. van Beek (Eds.), Sateli, B., "Semantic Management of Scholarly Literature: A Wiki-based Approach ", *The 27th Canadian Conference on Artificial Intelligence (Canadian AI 2014)*, vol. LNCS 8436, Montréal, Canada : Springer, pp. 387–392, 04/2014.

14. Witte, R., and B. Sateli, "Adding Natural Language Processing Support to your (Semantic) MediaWiki", *The 9th Semantic MediaWiki Conference (SMWCon Spring 2014)*, Montreal, Canada, 05/2014.

Source URL (retrieved on *2026-01-25 15:37*): https://www.semanticsoftware.info/semantic-assistants-wiki-nlp-showcase