

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

Adding Natural Language Processing Support to your (Semantic) MediaWiki

René Witte Bahar Sateli

Semantic Software Lab
Concordia University
Montréal, QC, Canada

SMWCon Spring 2014

Natural Language Processing (NLP)

- ▶ A branch of Artificial Intelligence
 - ▶ uses various techniques to process content written in natural language
- ▶ Multitude of NLP techniques
 - ▶ Named Entity Recognition
 - ▶ Question-Answering
 - ▶ Summarization
- ▶ Various NLP tools (e.g. [GATE](#), UIMA, ...) and APIs (e.g. OpenCalais, AlchemyAPI, ...)

Introduction

Natural Language Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

BBC News - Egypt crisis: Clashes in Cairo amid constitution row

Egypt crisis: Clashes in Cairo amid constitution row

Rival protesters have clashed outside the presidential palace in the Egyptian capital, Cairo, as unrest grows over a controversial draft constitution.

Stones were thrown and supporters of President Mohamed Morsi dismantled tents set up by anti-Morsi protesters.

Vice President Mahmoud Mekki has said a referendum on the draft will go ahead on 15 December despite the unrest.

But he indicated that changes could be made after the vote, saying the "door for dialogue" remained open.

He urged critics of the draft document to put their concerns in writing for future discussion.

The critics say the draft was rushed through parliament without proper consultation and that it does not do enough to protect political and religious freedoms and the rights of women.

The draft added to the anger generated by Mr Morsi passing a decree in late November which granted him wide-ranging new powers.

'Breakthrough'

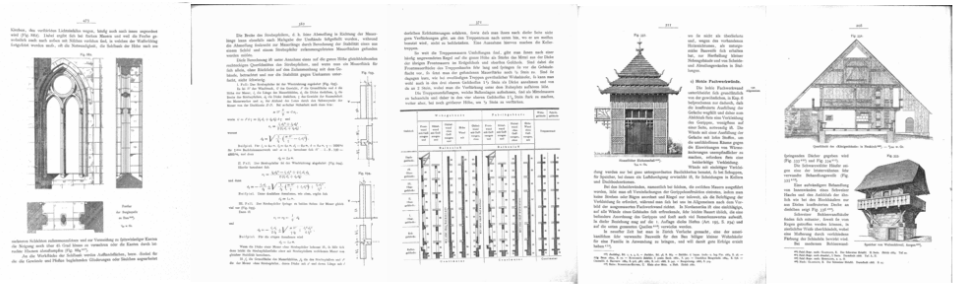
Egyptian Vice-President: 'Door open'

In a news conference broadcast live on state television, Mr Mekki said there was "real political will to pass the current period and respond to the demands of the public"

Entity Recognition Tags:

- ☐ Address
- ☒ Date
- ☐ FirstPerson
- ☒ JobTitle
- ☒ Location
- ☐ Lookup
- ☒ Organization
- ☐ Percent
- ☒ Person
- ☐ Sentence
- ☐ SpaceToken
- ☐ Split
- ☐ Title

Cultural Heritage Data Management



Large unstructured corpora

- Outdated terms, style of writing, huge amount, no categorization or assessment
- Comparing and evaluating with current content

Wiki Version of 19th-century documents

The Durm project

- Digitized version of a historical encyclopedia for architecture
- Provides automatic index generation



Navigation

- Hauptseite
- Durm-Projekt
- Zufälliger Artikel

Suche

Los Suche

Werkzeuge

- Was zeigt hierhin
- Verlinkte Seiten
- Hochladen
- Spezialseiten
- Druckversion
- Permanentlink

[Artikel](#) [Diskussion](#) [bearbeiten](#) [Versionen](#)

Wände aus Eisen und Stein. (Eisenfachwerkbau.): Schluß

in Kapitel 8: [Wände aus Eisen und Stein. \(Eisenfachwerkbau.\)](#); vorheriges Unterkapitel: [Durm:Bildung des Wandschlusses.](#) - [Inhaltsverzeichnis des Hefes](#)

Inhaltsverzeichnis [\[Verbergen\]](#)

- 1 Formale Behandlung. (232.)
- 2 Wertschätzung. (233.)

Seite 270 [\[Scan\]](#)

Formale Behandlung. (232.) [\[bearbeiten\]](#)

Von einer architektonischen Ausbildung des Eisenfachwerkbaues kann selbstverständlich nur die Rede sein, wenn das Eisengerippe ganz oder zum größten Teile unverhüllt gezeigt wird. Die Schwierigkeit derselben beruht, wie bei allen Eisenkonstruktionen, die in Verbindung mit Stein auszuführen sind, in der Magerkeit der mit Rücksicht auf Zweckmäßigkeit angeordneten und bemessenen Eisenteile, namentlich bei der Verwendung von Walzeisen und Blech. Auch die am besten durchgebildeten Eisenfachwerkbauten werden ein gewisses trockenes und hartes Aussehen nicht abstreifen können, da die Düntheit der Wände kräftige Laibungen der Öffnungen und deren belebende Schattenwirkung nicht zuläßt und alle Gliederungen naturgemäß an Fleischlosigkeit leiden. Wesentlich günstiger in Bezug auf die Schattenwirkung der Öffnungen sind wegen der tieferen Laibungen die als Hohlwände ausgeführten Eisenfachwerke, welche auf der Pariser Weltausstellung 1889 bei mehreren Bauwerken zur Anwendung



Scan der Originalseite 270

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

The first Wiki-NLP Experiments (ca. 2004)

Link to Discussion
Page for this Section

Link to Chapter
of Section

Section Heading
Main Table
of Content

Section Table
of Content

Full Text Search
New Page in
Original Document

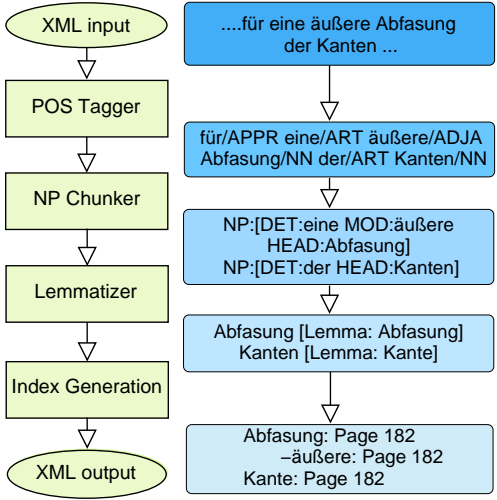
Link to Figures in
Original Document

Subsection
Heading

Preview of Scan
of Original Page
and Link to
High-Resolution
Version

Converting a Historical Architecture Encyclopedia into a Semantic Knowledge Base [WKKL10]

Back-of-the-Book Index Generation



| | | | |
|---------|------------|------------|-----------|
| artikel | diskussion | bearbeiten | versionen |
|---------|------------|------------|-----------|

Index

| | | | |
|---------------|-----|-----|-----|
| Abbinden | 137 | | |
| Abbrechen | 143 | | |
| Abdecken | 125 | | |
| Abdeckung | 131 | 108 | |
| Abfallstoffen | | | |
| oft | 129 | | |
| werdenden | 129 | | |
| lästig | 129 | | |
| Abfasung | 19 | 181 | 139 |
| ausgedehnte | 186 | | |
| angebrachte | 180 | | |
| äußere | 182 | | |
| äußerer | 182 | | |

Opportunities for NLP in Wikis

What could NLP provide for wikis?

- ▶ Detect various defects in its content
e.g., spelling mistakes, ambiguities, readability issues
- ▶ Extract entities that are relevant to a user's interest or context
e.g., extract all person names mentioned in the wiki
- ▶ Formally model the knowledge contained inside the wiki
e.g., generate Semantic MediaWiki markup from unstructured wiki text
[[hasType :: Enzyme | Xylanase]]
- ▶ Offer searching for content beyond keyword-based approaches
e.g., find all articles containing an enzyme name
- ▶ Generate its own content
e.g., create summaries from long articles
- ▶ Help structuring a wiki
e.g., suggest terms to link to Wikipedia

WikiSym 2007: Connecting Wikis and Natural Language Processing Systems [WG07]

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP
Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

General Architecture for Text Engineering (GATE)

Open Source Text Analytics

Mature framework, more than 10 years of development

Development team at University of Sheffield, UK (gate.ac.uk)

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP
Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

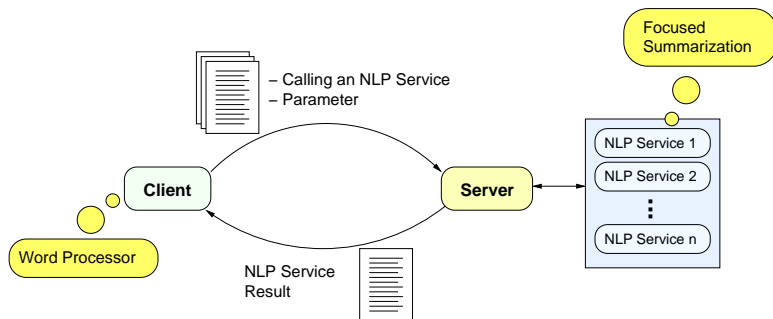
The screenshot displays the GATE Developer 6.0 interface. The main window shows a text document "text.txt_0001B" with the sentence "The field is updated." highlighted. Below the text, a table lists annotations with their types, sets, start/end positions, and IDs. The table is as follows:

| Type | Set | Start | End | Id | Description |
|---------------------|-----|-------|-----|----|---|
| Token | | 0 | 3 | 1 | [category=DT, kind=word, length=3, orth=upperInitial, string=The] |
| ReadabilityAnalysis | | 0 | 21 | 16 | [Explanation=The (sentence) has been detected as passive, and G |
| Token | | 4 | 9 | 3 | [category=NN, kind=word, length=5, orth=lowercase, string=field] |
| Token | | 10 | 12 | 5 | [category=VBZ, kind=word, length=2, orth=lowercase, string=is] |
| VG | | 10 | 20 | 14 | [tense=SimPre, type=FVG, voice=passive] |
| Token | | 13 | 20 | 7 | [category=VBN, kind=word, length=7, orth=lowercase, string=updated] |
| Token | | 20 | 21 | 8 | [category=., kind=punctuation, length=1, string=.] |

The interface also includes a left sidebar with a project tree, a top menu bar, and a right sidebar with a list of annotation sets and a document editor at the bottom.

Semantic Assistants

- ▶ Service-oriented Architecture (SOA) [WG08]
- ▶ Publishes various NLP pipelines as W3C Standard Web services
- ▶ Developed by Semantic Software Lab (<http://www.semanticsoftware.info>)
- ▶ Client plug-ins for OpenOffice, Eclipse, MediaWiki, Liferay, Android
- ▶ Open source framework (<http://www.semanticassistants.com>)



Introduction

Natural Language
Processing

Motivation
Semantic Assistants

The Wiki-NLP
Integration

Requirements Analysis
System Design
Implementation

Hands-On

Configuration
Information Extraction with
ANNIE
Semantic Queries

Application Examples

ReqWiki
Motivation
ReqWiki Design
ReqWiki Evaluation
IntelliGenWiki

Conclusions

References

Introduction

The Wiki-NLP Integration

Requirements Analysis

System Design

Implementation

Hands-On

Application Examples

Conclusions

References

Problem Statement

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

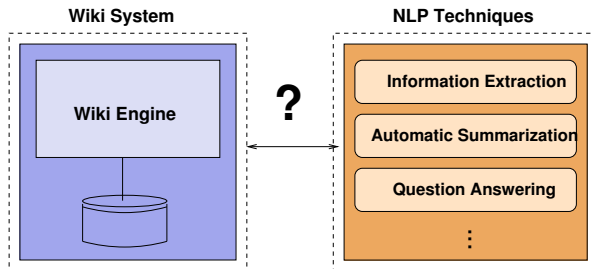
ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References



Solve common problems

- ▶ Wikis' Loose Structure
- ▶ Information Overload

Introduce new features

- ▶ Enable Human-AI collaboration
- ▶ Bring semantics to wiki content

Solution

Semantic Assistants Wiki-NLP Integration

- ▶ See [SW12] and <http://www.semanticsoftware.info/semantic-assistants-wiki-nlp>

Requirements Analysis

Three perspectives:

1. Wiki End-user Requirements

- ▶ From laypersons to technical employees
- ▶ Different cognitive backgrounds

2. Wiki Developer Requirements

- ▶ Engine developers
- ▶ Wiki administrators

3. System Requirements

Introduction

Natural Language
Processing

Motivation
Semantic Assistants

The Wiki-NLP
Integration

Requirements Analysis
System Design
Implementation

Hands-On
Configuration
Information Extraction with
ANNIE
Semantic Queries

Application Examples

ReqWiki
Motivation
ReqWiki Design
ReqWiki Evaluation
IntelliGenWiki

Conclusions

References

Requirements Analysis

NLP Service Independence

Run any NLP pipeline on wiki content

Seamless Integration

No context switching between wiki and the NLP framework

Easy Deployment

No major modification on the wiki engine

Collection-based Analysis

Multi-document analysis of wiki content

Other Requirements

Read Content from Wiki, Write Results to Wiki, Flexible Response Handling,
Proactive Service Execution, ...

Requirements Definitions

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

Seamless Integration

Using the Wiki-NLP integration must not largely deviate from the established usage patterns of wikis

Easy Deployment

Wiki-NLP integration must not require major changes to the wiki engine or to the means to access the wiki

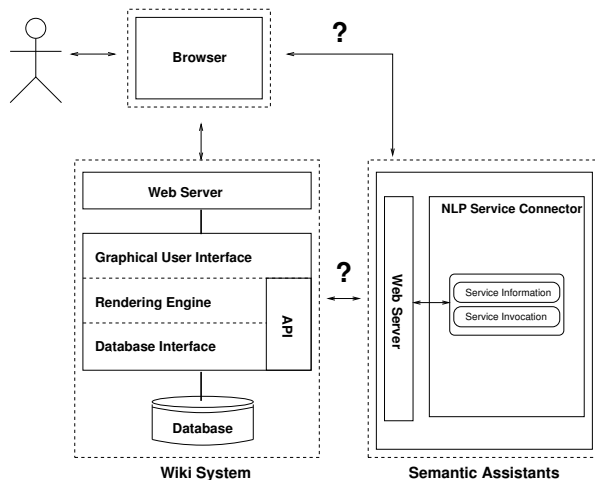
Collection-based Analysis

Users must be able to collect pages of interest and run an NLP service on the collection at once

Design Decisions

Design Questions

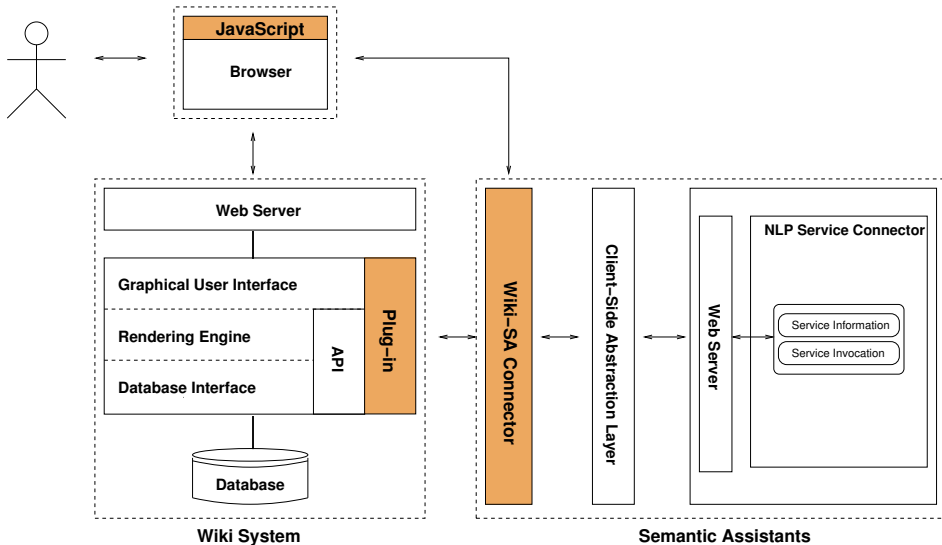
- Where should the Wiki-NLP integration be implemented?
- How can the system interactions be realized?



Design Decisions (II)

Collaborative approach

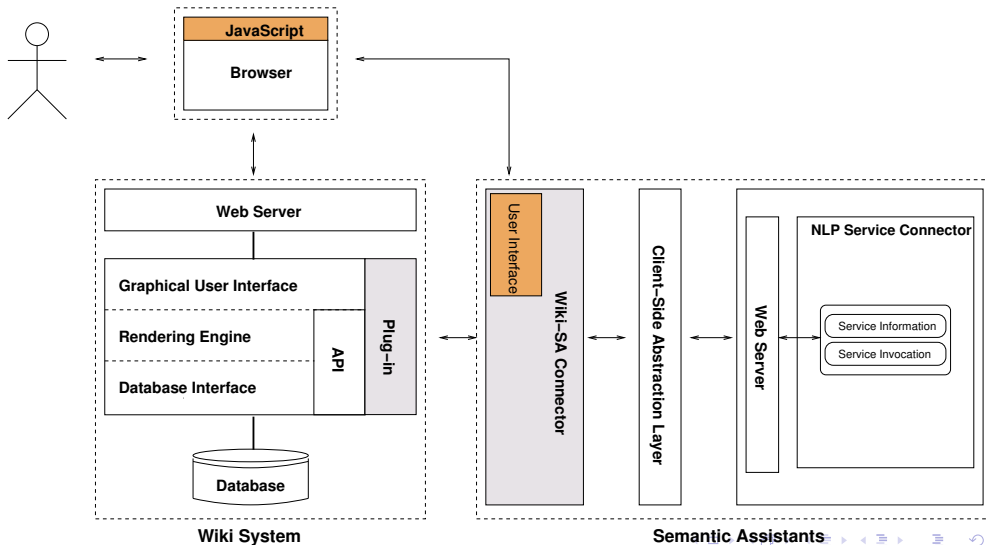
Division of responsibilities based on each system capabilities



User Interaction

Dynamic User Interface

Generating the user interface on-the-fly



Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP
Integration

Requirements Analysis

System Design
Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

Dynamic User Interface

Requirements

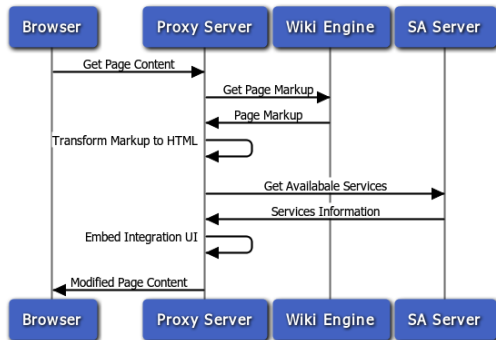
- ▶ Seamless Integration
- ▶ Easy Deployment

Solution

Proxy the wiki page request and embed the user interface on-the-fly

Advantages

- ▶ Wiki-independent
- ▶ No deployment required
- ▶ Seamless environment



Dynamic User Interface

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

Requirement

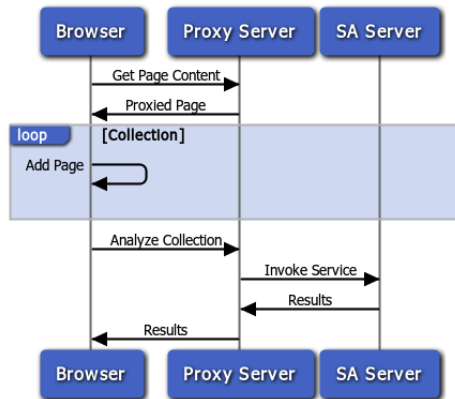
Collection-based Analysis

Solution

Preserve user's selection in the browser before the actual service invocation (e.g., cookies)

Advantages

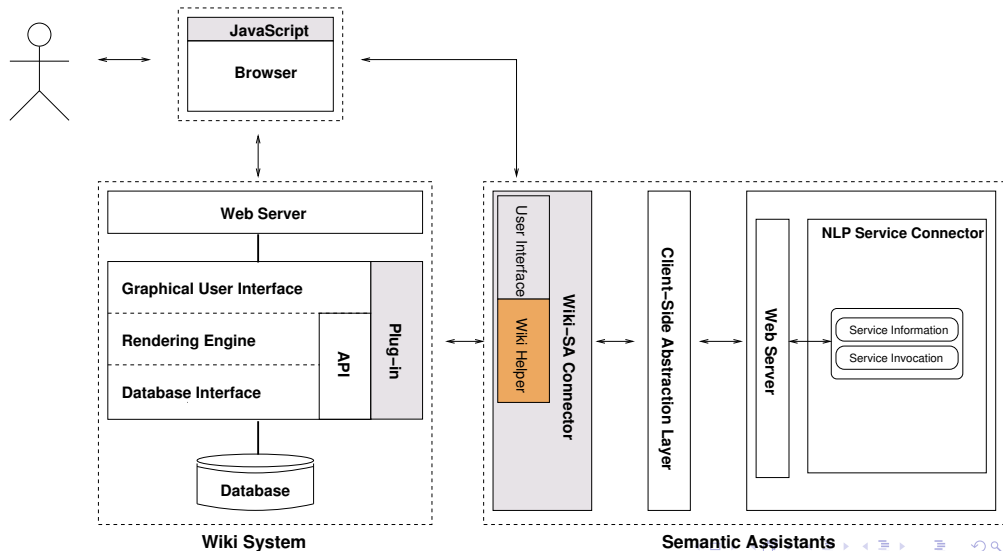
- ▶ Multi-document Analysis
- ▶ Increased efficiency



Wiki Communication

Wiki Bot Frameworks

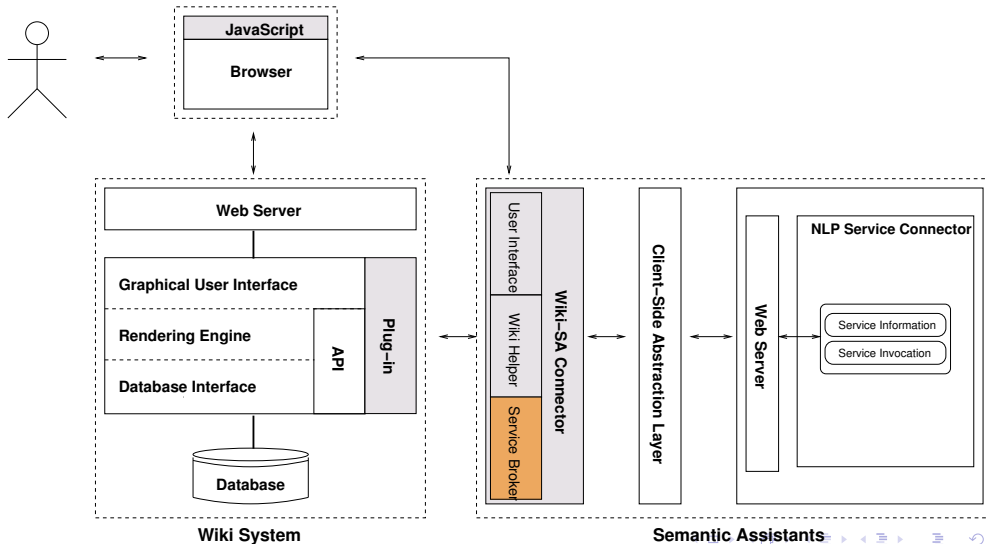
Accessing wiki content via its API with a higher level of abstraction



Service Invocation

Service Broker Component

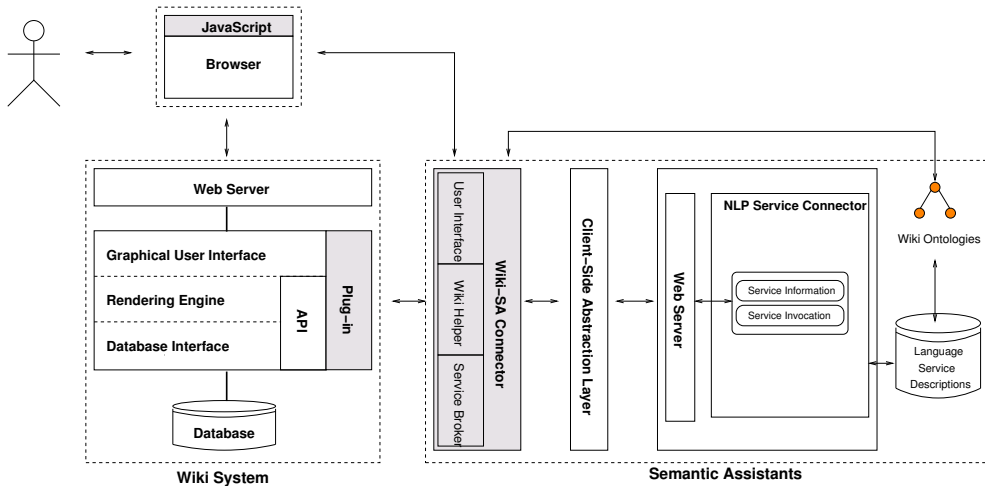
Delegating service invocation requests to the Semantic Assistants server



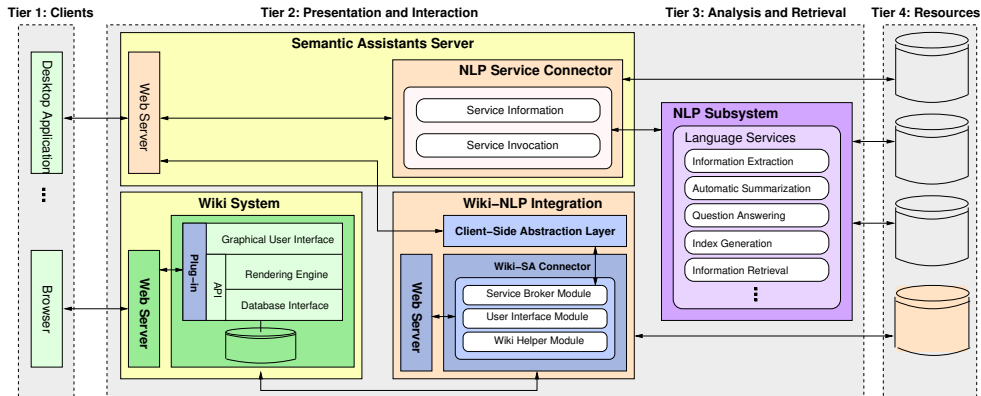
Wiki Independence

Semantics-based Architecture

A repository of formalized descriptions of wikis written in OWL



Merged System Architecture



The new Wiki-NLP integration module in the Semantic Assistants architecture

Transformation of Results

From the Semantic Assistants server response to wiki markup

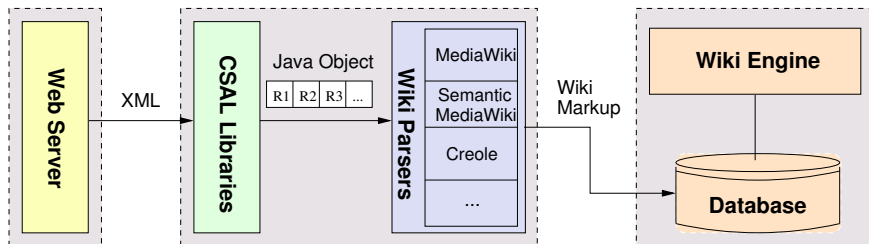
Sample Service Invocation Response

```
<saResponse>
  <annotation type="Location" annotationSet="Annotation" isBoundless="false">
    <document url="http://localhost/wiki/sample_page">
      <annotationInstance content="Canada" start="16" end="22">
        <feature name="locType" value="country"/>
      </annotationInstance>
    </document>
  </annotation>
</saResponse>
```

Semantic Assistants

Wiki-SA Connector

Wiki System



Presentation of Results

Templating mechanism

Separating data model from its presentation

```

1 { | class="wikitable" style="height:50px"
2 | width="200" | Content
3 | width="80" | Type
4 | width="50" style="text-align: center;" | Start
5 | width="50" style="text-align: center;" | End
6 | Features
7 |− valign="top" | {{{content}}} | style="text-align: center;" | ((Property:{{{type}}}|{{{type}}})) |
   style="text-align: center;" | {{{start}}} | style="text-align: center;" | {{{end}}} | {{{
   features}}}
8 |}

```

```

1 {{SemAssist−TableRow| content= Elizabeth Middleton | type=Person | start = 236 | end = 255 | features =
   gender:female}}

```

| Content | Type | Start | End | Features |
|---------------------|--------|-------|-----|------------------|
| Elizabeth Middleton | Person | 236 | 255 | ■ gender: female |

The NLP Interface

The NLP user interface can now offer various text mining services

- Dynamically-generated interface
- Customizing services at runtime
- Collection-based Analysis

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP
Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

Text Mining Assistants inside the wiki

PubMed:20709852

Title: Characterization of a cellobiohydrolase (MoCel6A) produced by *Magnaporthe oryzae*.

Authors: Takahashi M, Takahashi H, Nakano Y, Konishi T, Terauchi R, Takeda T.

Institute: Iwate Biotechnology Research Center, Kitakami, Iwate, Japan.

PMID: 20709852

Received on March 10, 2010. Accepted on July 30, 2010.

Full Text [edit]

Abstract

Available Assistants Results Target Global Settings Console

Step 1. Select the service your wish to execute on your collection.
Once you add this page to your collection, you can continue browsing as your collection is saved.

Available Assistants Select a service

Runtime Parameters Select a service

Collection

mycoMINE

IR Information Extractor

Information Extractor

OrganismTagger

Add

Clear

This page was last modified on 6 November 2012, at 23:21. This page has been accessed 4 times. Privacy policy About

IntelliGenWiki Disclaimers

Available Assistants Results Target Global Settings Console

Step 1. Select the service your wish to execute on your collection.
Once you add this page to your collection, you can continue browsing as your collection is saved.

Available Assistants Select a service

Runtime Parameters Select a service

Collection

mycoMINE

IR Information Extractor

Information Extractor

OrganismTagger

Add

Clear

Done

Implementation

Information Extraction with ANNIE

RegWiki

Motivation

RegWiki Design

RegWiki Evaluation



navigation

- [Main page](#)
- [Recent changes](#)
- [Help](#)

documents

- Vision
- Use Case
- Supplementary Specification

toolbox

- [What links here](#)
- [Related changes](#)
- [Upload file](#)
- [Special pages](#)
- [Printable version](#)
- [Permanent link](#)
- [Semantic Assistants](#)
- [Browse properties](#)

search

[page](#)
[discussion](#)
[edit](#)
[history](#)
[delete](#)
[move](#)
[protect](#)
[watch](#)
[refresh](#)

BP Oil Spill

On 20 April 2010, a semi-submersible exploratory offshore drilling rig in Mexico exploded after a blowout. The US Government named British Petroleum, a London headquartered company, the responsible party. Eventually, the company stopped the flow of crude by shutting off pipes that were funneling some of the oil to ships on the surface on 14 July.

Following the oil spill the company stock fell in half and there were speculations in the press, guided by the commentary of Fred Lucas, that there would be a takeover of the company, focusing on possible bids from Exxon and Shell at a presumed price of £88 billion.

[Information Extractor on BP Oil Spill \(View\)](#)

| Content | Type | Start | End | Features |
|----------------------|--------------|-------|-----|-----------------------|
| 20 April 2010 | Date | 3 | 16 | ■ kind: date |
| 14 July | Date | 340 | 347 | ■ kind: date |
| No annotations found | Address | | | |
| £88 billion | Money | 603 | 614 | ■ kind: number |
| Fred Lucas | Person | 474 | 484 | ■ gender: male |
| US Government | Organization | 111 | 124 | ■ orgType: government |
| British Petroleum | Organization | 131 | 148 | ■ orgType: company |
| Exxon | Organization | 564 | 569 | ■ orgType: company |
| Shell | Organization | 574 | 579 | ■ orgType: company |
| Mexico | Location | 74 | 80 | ■ locType: city |
| London | Location | 152 | 158 | ■ locType: city |
| No annotations found | Percent | | | |

Introduction

Natural Language
Processing

Motivation
Semantic Assistants

The Wiki-NLP
Integration

Requirements Analysis
System Design
Implementation

Hands-On

Configuration
Information Extraction with
ANNIE
Semantic Queries

Application Examples

ReqWiki
Motivation
ReqWiki Design
ReqWiki Evaluation
IntelliGenWiki

Conclusions

References

Introduction

The Wiki-NLP Integration

Hands-On

Configuration

Information Extraction with ANNIE

Semantic Queries

Application Examples

Conclusions

References

Wiki-NLP: Initial Configuration

We recommend you use *Firefox* for the hands-on exercises.

1. Browse to <http://hostname/wikinlp>
2. Login with the following credentials
 - ▶ Username: *********
 - ▶ Password: *********
3. Click on the “Semantic Assistants” link on the left
4. Configure the Wiki-NLP integration with the same credentials as above

Semantic Assistants

Before you proceed, you have to specify the followings: (all fields are required)

Wiki Engine

MediaWiki-1.16

Wiki Address

Username

Password

Semantic Assistants Server

minion.cs.concordia.ca:8879

Save

Semantic Software Lab

Wiki-NLP: Information Extraction

1. Who's your favorite celebrity or scholar? Pick one.
2. Create a sample page with his/her mini-biography (Copy-Paste from Wikipedia)
3. Click on the "Semantic Assistants" link on the left
4. Choose a service from the "Available Assistants"
5. Add the page to your "Collection"
6. Go to the "Results Target" tab, investigate the options and click "Run Service"
7. When finished, click on the "Page" tab on top to refresh the page and see the results

Available Assistants

Results Target

Global Settings

Console

Step 1. Select the service your wish to execute on your collection. Once you add this page to your collection, you can continue browsing as your collection is saved.

Available Assistants

Person and Location Extractor ▾

Collection

Add

Clear

Runtime Parameters
This service has no runtime parameter.

Wiki-NLP: Semantic Queries

1. In the page you created, type the following query to find all the “Person” entities:

```
{{#ask: [[hasType::Person]]
|?Person= Entities Found
|format=table
|headers=plain
|default=No pages found!
|mainlabel=Page Name
}}
```

(Copy-paste from the “Sample Semantic Query” on the Main Page)

2. Discover all other persons mentioned in the wiki!

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP
Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

Introduction

The Wiki-NLP Integration

Hands-On

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

Requirements Engineering (RE) Tool Support

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

Wiki Systems for Software Requirements Engineering

- ▶ Collaborative environment for elicitation
 - ▶ Alternative to existing tools, better support, cost
 - ▶ Discussion media: sharing, traceability, guidelines
- ▶ Moderator
 - ▶ Channel stakeholders' contribution: templates, assignments, overviews
 - ▶ Indicators analysis for conflicts detections
- ▶ Requirements refinements
 - ▶ Reviews
 - ▶ Extension to capture additional Requirements (NFR)

Requirements in Practice

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

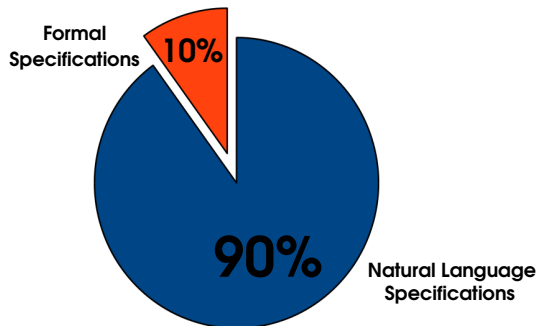
ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References



Industry distribution of SRS [MFI04]

Issues

- ▶ Natural Language SRS are fragile & difficult to maintain
- ▶ Limited tool support for automatic analysis

RE Tool Support

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

Wiki Improvements for RE

- ▶ Existing Wikis need further support for RE (added using plug-in architecture)
 - ▶ Improved page **linking**: automation, better contribution (traceability)
 - ▶ Lack of **integration**: connectors for other tools (e.g., CVS/Subversion) or Wiki integration into software development tools (Trac, Redmine)
 - ▶ Lack of **semantics**: add semantic layer (using ontologies), e.g., Semantic MediaWiki (SMW)

ReqWiki: A Semantic System for Collaborative Requirements Engineering

Open Source Project at Concordia's Semantic Software Lab

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References



navigation

- [Main page](#)
- [Recent changes](#)
- [Help](#)

documents

- [Vision](#)
- [Use Case](#)
- [Supplementary Specification](#)

toolbox

- [What links here](#)
- [Related changes](#)
- [Special pages](#)
- [Printable version](#)
- [Permanent link](#)
- [Semantic Assistants](#)
- [Browse properties](#)

search

[Log in](#)

[page](#)

[discussion](#)

[view source](#)

[history](#)

Main Page

Wiki-based Software Requirements Engineering

Wiki FAQ

- [MediaWiki User Guide](#)
- [MediaWiki Formatting Guide](#)

ReqWiki System

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

Goals

- ▶ Wiki-based user interface for easy collaboration
- ▶ Creation of structured, semantic markup
- ▶ Standard data format for web integration
- ▶ Integrated semantic support, like traceability link management

Platform

- ▶ MediaWiki Engine
- ▶ Semantic MediaWiki extension
- ▶ Semantic Forms, Templates
- ▶ Semantic Assistants for NLP support

ReqWiki System Development

Transforming the 20th-century UP templates into a Semantic Wiki

- **Idea:** start from original *Word* document templates (which are based on the [Rational] Unified Process ([R]UP) industry standard)
- Transform them into a modern, Semantic Wiki version
- See [SAW13] and <http://www.semanticsoftware.info/reqwiki>

2.1. Problem Statement

| | |
|---------------------------------------|---|
| The problem of | <i>Difficulty comparing nutrition values of similar aliment-products</i> |
| Affects | <i>System Users</i> |
| The impact of which is | <ul style="list-style-type: none"> - <i>Inaccurate or raw-estimated food consumptions</i> - <i>Unable to make balanced food purchases to optimize healthy eating habits.</i> - <i>Poor nutrition</i> |
| A successful solution would be | <ul style="list-style-type: none"> - <i>Clarify nutrition-fact-labels</i> - <i>Food classification</i> |

Introduction

- Natural Language Processing
- Motivation
- Semantic Assistants

The Wiki-NLP
Integration

- Requirements Analysis
- System Design
- Implementation

Hands-On

- Configuration
- Information Extraction with ANNIE
- Semantic Queries

Application Examples

- ReqWiki
- Motivation
- ReqWiki Design
- ReqWiki Evaluation
- IntelliGenWiki

Conclusions

References

Top-level
pages

Semantic
Assistants
Plug-in

navigation

- Main page
- Recent changes
- Help

documents

- Vision
- Use Case
- Supplementary Specification

toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link
- Semantic Assistants
- Browse properties

search

Go

Search

page

discussion

edit

history

delete

move

protect

watch

refresh

Sysop my talk my preferences my watchlist my contributions log out

Vision

Contents [show]

1. Introduction

State purpose of the vision document and describe the purpose of the project/software solution

2.Positioning

Provide a statement summarizing the problem being solved by this project.

2.1 Problem Statement

Create a stakeholder
Create a problem

| The problem of | StakeHolder | The Impact of Which is | A Successful Solution would be |
|---|--------------|--|---|
| Difficulty comparing nutrition values of similar aliment-products | System Users | <ul style="list-style-type: none">- Inaccurate or raw-estimated food consumptions- Unable to make balanced food purchases to optimize healthy eating habits.- Poor nutrition | <ul style="list-style-type: none">- Clarify nutrition-fact-labels- Food classification |

2.2 Product Position Statement

Links to semantic forms

Dynamic tables generated from semantic queries

Using “Semantic Forms” Extension

MediaWiki Extension “Semantic Forms” allows to create forms that generate [semantic markup](#)

Natural Language
Processing

Information Extraction with ANNIE

RegWiki

Motivation

RegWiki Design

ReqWiki Evaluation

[page](#) [discussion](#) [edit with form](#) [edit](#) [history](#) [delete](#) [move](#) [protect](#) [watch](#) [refresh](#)

Edit FormProblem: Difficulty comparing nutrition values of similar aliment-products

Affects:


Impact:

- Inaccurate or raw-estimated food consumptions
- Unable to make balanced food purchases to optimize healthy eating habits.
- Poor nutrition

Successful Solution:

- Clarify nutrition-fact-labels
- Food classification

User Interface: Semantic Forms



navigation

- [Main page](#)
- [Recent changes](#)
- [Help](#)

documents

- [Vision](#)
- [Use Case](#)

[page](#) [discussion](#) [edit with form](#) [edit](#) [history](#) [delete](#) [move](#)

Edit FormProblem: PoorNutrition

Affects:

Impact:

Short & long term health:

- * Becoming over/under weight.
- * Weakened immune system.
- * Lack of energy.
- * Serious problems (Hypertension, Diabetes, Cholesterol, Gout, Osteoporosis)

Successful Solution:

Combination of:

- * Balanced diet.
- * Track food intake.
- * Take vitamins & supplements.

Features

- ▶ Structures content at finer granularities
- ▶ Auto-completion for referencing entities (e.g., actors)
- ▶ Automatically generates RDF markup for linking & querying



Semantic Markup

The Semantic Forms automatically generate semantic markup (RDF format)

special page

[Browse wiki](#)

Difficulty comparing nutrition values of similar aliment-products

| | |
|----------------------------------|--|
| BelongsTo | System Users +  |
| HasImpact | - Inaccurate or raw-estimated food consumptions - Unable to make balanced food purchases to optimize healthy eating habits. - Poor nutrition |
| HasSolution | - Clarify nutrition-fact-labels - Food classification |
| Modification date | 29 November 2011 06:44:28 +  |
| Categories | Problem |
| hide properties that link here | |
| No properties link to this page. | |

ReqWiki System

Templates

To make the semantic markup readable by end users, **templates** are used:

page

discussion

edit with form

edit

history

delete

move

protect

watch

refresh

Difficulty comparing nutrition values of similar aliment-products

| | |
|--------------------------------|--|
| Affects | System Users |
| The Impact of which is | <ul style="list-style-type: none">- Inaccurate or raw-estimated food consumptions- Unable to make balanced food purchases to optimize healthy eating habits.- Poor nutrition |
| A successful solution would be | <ul style="list-style-type: none">- Clarify nutrition-fact-labels- Food classification |

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

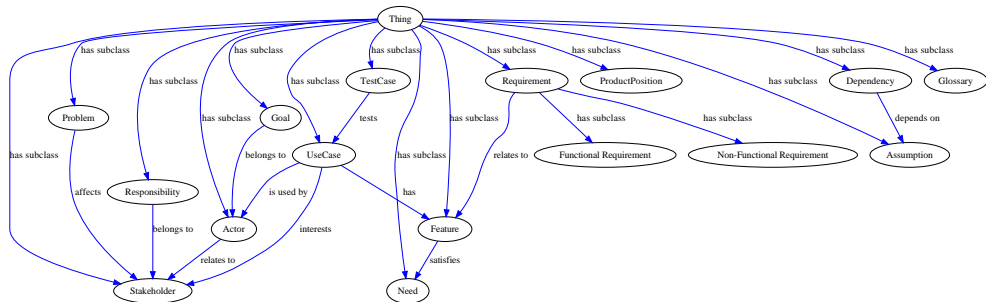
References

Semantic Model for SRS

Ontology

Content is structured using a Semantic Web ontology:

- Formally describe software artifacts and their components
- Used to model, connect and query SRS statements with ontology concepts

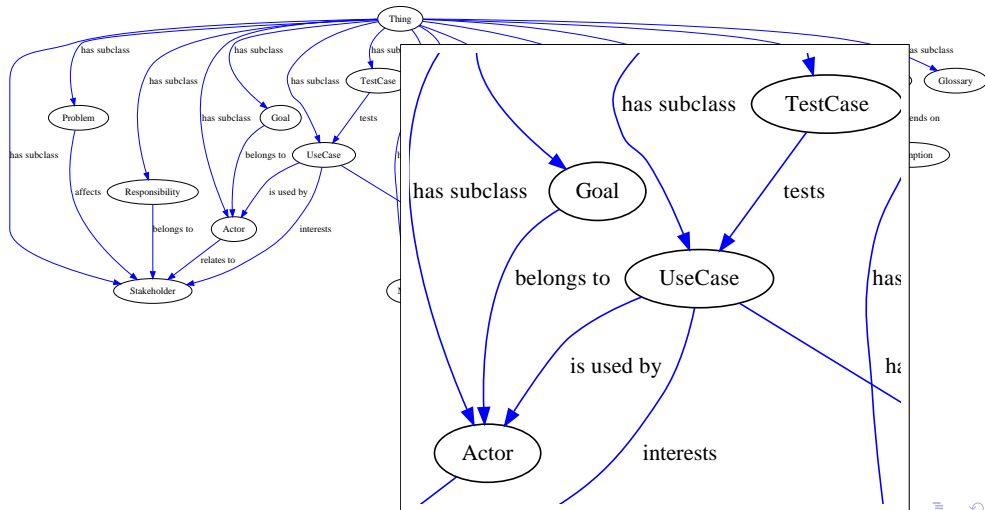


Semantic Model for SRS

Ontology

Content is structured using a Semantic Web ontology:

- Formally describe software artifacts and their components
- Used to model, connect and query SRS statements with ontology concepts



ReqWiki System

Embedded Queries

- Query for all stakeholders in the wiki:

```
1  {{#ask: [[<page>]]| [<param1> | <param2> ..] }}
```

2

3 Example:

4

```
5  {{#ask: [[Category:StakeHolder]]
6  | mainlabel=StakeHolder
7  | ?HasDescription= Description
8  | format= table
9  }}
```

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

[edit]

Create a responsibility



| StakeHolder | Description |
|---------------------|---|
| Developer | The person who is developing the application |
| Student | The person who is using the application |
| Teacher's Assistant | The person who is supervising a tutorial or lab |

ReqWiki: Traceability

Automatic Traceability Support

- ▶ Traceability is concerned with interrelating various software artifacts
- ▶ Manually cross-referencing documents is time-consuming and error-prone
- ▶ Exploiting the semantic metadata in ReqWiki
- ▶ Supported traceability links types:
 1. Revision links
 2. Semantic links
 3. Query-based links

User Needs versus Features

|  Need |  Feature |
|--|---|
| Modify policy detail information | Alter policy information |
| Modify policy detail information | Query the status of policy information |
| Alteration of unit link product | Input conversion to unit investment |
| Alteration of unit link product | Modify conversion to unit investment |
| Alteration of unit link product | Query unit investment |
| Alteration of unit link product | Query unit price |

```
{ {#ask: [[Category:Features]]  
| ?BelongsTo=Need  
| ? = Feature  
| format= table  
}}
```

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

ReqWiki System: Automatic Traceability Link Support

Traceability Links

- ▶ Traceability Links can now all be automatically generated
- ▶ Tables are created with a combination of Queries & Templates
- ▶ Dead links (e.g., after deleting an entity) can also be automatically detected
- ▶ Like all other semantic metadata, RDF can be exported from the Wiki and consumed by other tools

User Needs versus Features

| Need | Feature |
|------------------|----------------|
| Food Consumption | Nutrition Fact |

Features versus Use Cases

| UseCases | Features |
|------------------------|----------------|
| Collect Aliment-Intake | Nutrition Fact |

Features versus Supplementary Requirements

| Features | Functional Requirements |
|----------------|-----------------------------|
| Nutrition Fact | Food Intake |
| Features | Non Functional Requirements |
| Nutrition Fact | Mean Time |
| Features | Supplementary Requirements |
| Nutrition Fact | Food Intake |
| Nutrition Fact | Mean Time |

Use Cases to Test Cases

| UseCase | TestCases |
|------------------------|------------------|
| Collect Aliment-Intake | Aliment Testcase |

ReqWiki: Content Analysis

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

Motivation

Most requirements are expressed in natural-language form. . .

Example Requirement

“The system shall configure the frame-rate as specified by the camera provided during initialization & the frame-rate shall be capable of being automatically modified based on bandwidth.”

Natural Language Processing

. . . but these requirements are prone to expensive-to-find defects.

Solution: use Natural Language Processing (NLP) to help users (see Requirements Quality Assurance, [vL09])

Text Mining Assistants

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

Natural-Language Requirements Defects

SRS defects addressable with ReqWiki [Lap09, vL09]:

- ▶ Spelling & Grammar
- ▶ Incompleteness
- ▶ Contradiction
- ▶ Ambiguity
- ▶ Unmeasurability
- ▶ Noise
- ▶ Over Specification
- ▶ Inadequacy
- ▶ Unintelligibility
- ▶ Poor Structuring
- ▶ Unfeasibility
- ▶ Options
- ▶ ...

Clear

Plug-in Dialog

NLP
Support

- ▶ Various NLP services

Available Assistants

Runtime Parameters

Select a service

Select a service

Information Extractor

Writing Quality

English Durm Indexer

Requirements QA Defects

Requirements QA Stats

Readability Metric Stats

ReadabilityMetrics

Person and Location Extractor

- ▶ SRS defects addressable
 - ▶ Spelling & Grammar
 - ▶ Incompleteness
 - ▶ Ambiguity
 - ▶ Poor Structuring
 - ▶ Passive voice
 - ▶ ...
- ▶ Automatically index the SRS
 - ▶ back-of-the-book style
 - ▶ complement the glossary
 - ▶ helping domain analysis

UC/Manage Tasks

| | |
|------------------------------|--|
| Description | The manager receives a customer service request. The manager directs the operation for creating, updating, deleting and querying a task. Some operations use either the automatic or manual task assignment functionality that were defined in the Supplementary Specification document. |
| Level | user-goal |
| Primary Actor | A / Manager |
| StakeHolders | Manager, Senior technician, Junior technician |
| Pre-Conditions | The manager must be identified and authenticated in the application |
| Success end condition | The task is created and assigned to the technicians with status Assigned. The tasks is updated and assigned to the technicians with status Assigned. The task is queried. The task is deleted. |
| Failure end condition | The task is created with status Submitted. |
| Features | Manage Task |

Writing Quality on UC/Manage Tasks [\(View\)](#)

| Content | Type | Start | End | Features |
|--------------|------|-------|-----|---|
| were defined | AtD | 236 | 248 | <ul style="list-style-type: none"> problem: Passive voice suggestion: - |
| must be | AtD | 434 | 441 | <ul style="list-style-type: none"> problem: Passive voice suggestion: - |
| is created | AtD | 521 | 531 | <ul style="list-style-type: none"> problem: Passive voice suggestion: - |
| The tasks is | AtD | 587 | 599 | <ul style="list-style-type: none"> problem: Subject Verb Agreement suggestion: The tasks are, The task is |

Index Generation (through Noun Phrase Identification)

Automatically created “back-of-the-book index”



navigation

- [Main page](#)
- [Recent changes](#)
- [Help](#)

documents

- [Vision](#)
- [Use Case](#)
- [Supplementary Specification](#)

toolbox

- [What links here](#)
- [Related changes](#)
- [Upload file](#)
- [Special pages](#)
- [Printable link](#)
- [Permanent link](#)
- [Semantic Assistants](#)
- [Browse properties](#)

[page](#)

[discussion](#)

[edit](#)

[history](#)

[delete](#)

[move](#)

[watch](#)

[refresh](#)

English Durm Indexer

| | |
|--------------|--|
| activity | 1. Introduction |
| amount | 1. Introduction |
| appendix | 9. Glossary |
| application | 1. Introduction |
| many | 9. Glossary |
| other | 1. Introduction |
| availability | 9. Glossary |
| belongsto | 1. Actor-Goal List |
| Goal | 1. Actor-Goal List |
| bulk | 9. Glossary |
| capability | 9. Glossary |
| capture | Supplementary Specification 9. Glossary |
| case | use 9. Glossary |
| category | 9. Glossary 1. Actor-Goal List 1. Introduction |
| cycle | task 1. Introduction |
| data | 9. Glossary |
| date | 9. Glossary |

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

User Study I – Effectiveness

Research Question

Can text mining assistants help to improve requirements specifications?

Methodology

User study with two software engineering classes at Concordia University

- ▶ **Goal:** Identifying defects in manual vs. NLP-assisted requirements specifications
- ▶ **NLP Services:** Spell checking, Readability Analysis, Passive Voice Detection, ...
- ▶ **Measure:** Average number of defects found in the two assignment revisions
- ▶ **Method:** Comparison of manual vs. NLP-assisted quality assurance

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

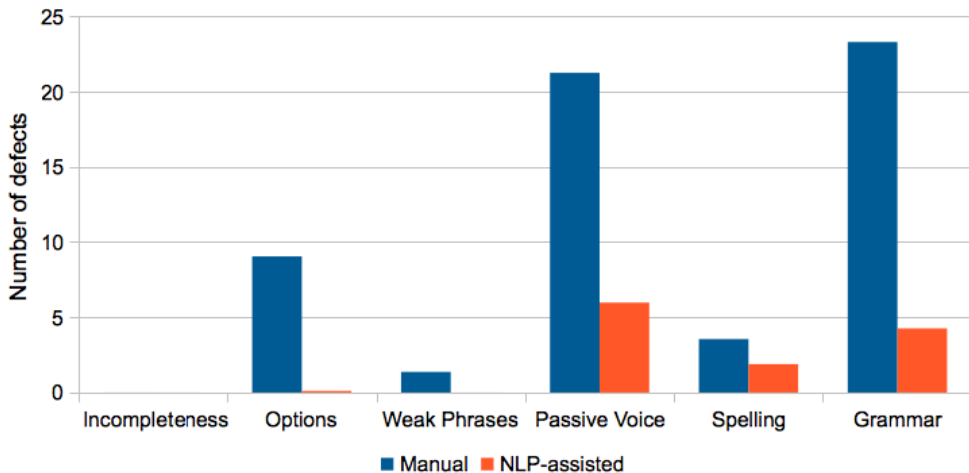
ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

User Study I – Effectiveness: Results



Conclusion

ReqWiki NLP capabilities were indeed effective to significantly reduce SRS defects.

User Study II – Usability

Research Question

How much NLP background do users need in order to use semantic capabilities?

Methodology

Same scenario as User Study I; Anonymized questionnaire asking participants about:

1. Their proficiency level in NLP
2. ReqWiki ease-of-use

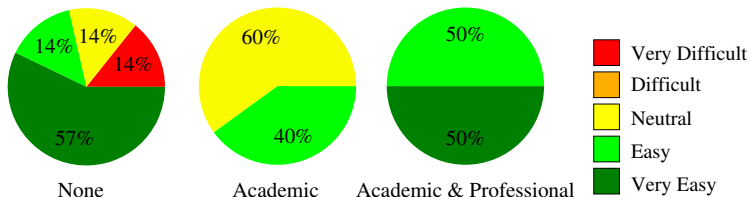
What is your level of experience in the area of Natural Language Processing?

Choose one of the following answers

- ☐ Previous academic experience (e.g., you have taken related courses)
- ☐ Previous industrial experience (e.g., you have worked in this area)
- ☐ Both academic and industrial experience
- ☐ None

User Study II – Usability: Results

Students' Questionnaire Feedback



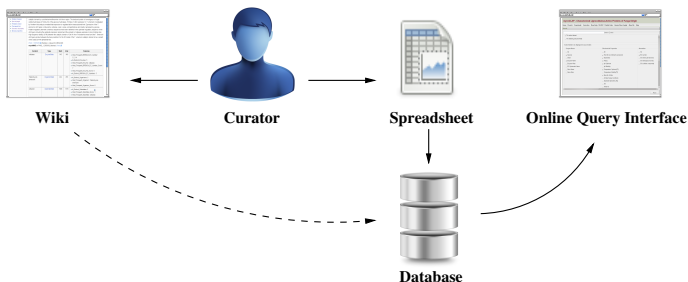
Conclusion

Concrete NLP background is *not* required to make use of sophisticated semantic support provided in ReqWiki.

Wikis as Knowledge Management Platforms

► Biomedical Literature Curation

- Finding and extracting relevant knowledge from the domain literature
- Manually refining and updating bioinformatics databases



► Manual literature curation is...

- **Expensive** → requires domain experts
- **Labour-intensive** → ever growing amount of scientific publications
- **Error-prone** → critical knowledge can be easily missed

IntelliGenWiki

- ▶ An intelligent semantic wiki for life sciences [SMB⁺12]
- ▶ Integrated bio-related NLP services for literature curation
- ▶ Offers basic semantic entity retrieval
- ▶ See <http://www.semanticsoftware.info/intelligenwiki>

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

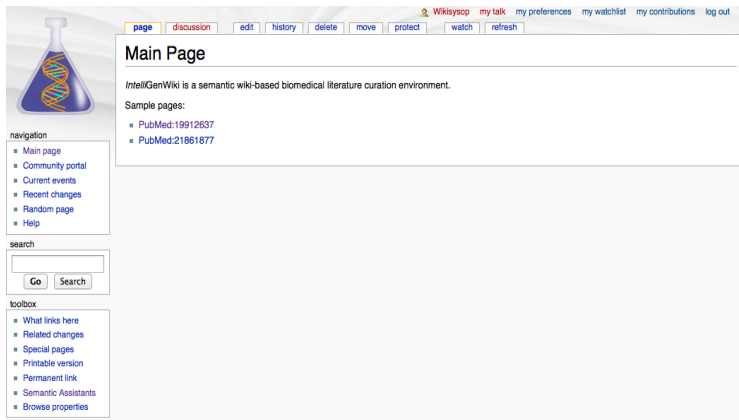
ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References



Wikisop my talk my preferences my watchlist my contributions log out

page discussion edit history delete move protect watch refresh

Main Page

IntelliGenWiki is a semantic wiki-based biomedical literature curation environment.

Sample pages:

- PubMed:19912637
- PubMed:21861877

navigation

- Main page
- Community portal
- Current events
- Recent changes
- Random page
- Help

search

Go Search

toolbox

- What links here
- Related changes
- Special pages
- Printable version
- Permanent link
- Semantic Assistants
- Browse properties

IntelliGenWiki: User Interface

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation


ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References



navigation

- [Main page](#)
- [Community portal](#)
- [Current events](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)

search

Go Search

toolbox

- [What links here](#)
- [Related changes](#)
- [Special pages](#)
- [Printable version](#)
- [Permanent link](#)
- [Semantic Assistants](#)
- [Browse properties](#)

PubMed:20709852 - IntelliGenWiki - Mozilla Firefox

[File](#)
[Edit](#)
[View](#)
[History](#)
[Bookmarks](#)
[Tools](#)
[Help](#)

PubMed:20709852 - IntelliGenWiki

+

[Wikisypop](#)
[my talk](#)
[my preferences](#)
[my watchlist](#)
[my contributions](#)
[log out](#)

[page](#)
[discussion](#)
[edit](#)
[history](#)
[delete](#)
[move](#)
[protect](#)
[watch](#)
[refresh](#)

PubMed:20709852

Title: Characterization of a cellobiohydrolase (MoCel6A) produced by *Magnaporthe oryzae*.

Authors: Takahashi M, Takahashi H, Nakano Y, Konishi T, Terauchi R, Takeda T.

Institute: Iwate Biotechnology Research Center, Kitakami, Iwate, Japan.

PMID: 20709852

Received on March 10, 2010. Accepted on July 30, 2010.

Full Text

[\[edit\]](#)

Abstract

Three GH-6 family cellobiohydrolases are expected in the genome of *Magnaporthe grisea* based on the complete genome sequence. Here, we demonstrate the properties, kinetics, and substrate specificities of a *Magnaporthe oryzae* GH-6 family cellobiohydrolase (MoCel6A). In addition, the effect of cellobiose on MoCel6A activity was also investigated. MoCel6A contiguously fused to a histidine tag was overexpressed in *M. oryzae* and purified by affinity chromatography. MoCel6A showed higher hydrolytic activities on phosphoric acid-swollen cellulose (PSC), β -glucan, and cellooligosaccharide derivatives than on cellulose, of which the best substrates were cellooligosaccharides. A tandemly aligned cellulose binding domain (CBD) at the N terminus caused increased activity on cellulose and PSC, whereas deletion of the CBD (catalytic domain only) showed decreased activity on cellulose. MoCel6A hydrolysis of cellooligosaccharides and sulforhodamine-conjugated cellooligosaccharides was not inhibited by exogenously adding cellobiose up to 438 mM, which, rather, enhanced activity, whereas a GH-7 family cellobiohydrolase from *M. oryzae* (MoCel7A) was severely inhibited by more than 29 mM cellobiose. Furthermore, we assessed the effects of cellobiose on hydrolytic activities using MoCel6A and *Trichoderma reesei* cellobiohydrolase (TrCel6A), which were prepared in *Aspergillus oryzae*. MoCel6A showed increased hydrolysis of cellopentaose used as a substrate in the presence of 292 mM cellobiose at pH 4.5 and pH 6.0, and enhanced activity disappeared at pH 9.0. In contrast, TrCel6A exhibited slightly increased hydrolysis at pH 4.5, and hydrolysis was severely inhibited at pH 9.0. These results suggest that enhancement or inhibition of hydrolytic activities by

Bio-NLP: Information Extraction

Introduction

Natural Language
Processing
Motivation
Semantic Assistants

The Wiki-NLP Integration

Requirements Analysis
System Design
Implementation

Hands-On

Configuration
Information Extraction with
ANNIE
Semantic Queries

Application Examples

ReqWiki
Motivation
ReqWiki Design
ReqWiki Evaluation
IntelliGenWiki

Conclusions

References

- ▶ Automatically extracting knowledge from papers

- ▶ Various IE services

- ▶ OrganismTagger
- ▶ Open Mutation Miner
- ▶ Idots

- ▶ Enrichment of literature content with semantic markup

Example:

`[[hasType::Enzyme|cellobiohydrolase]]`

severely inhibited at pH 9.0. These results suggest that enhancement or inhibition of hydrolytic activities by cellobiose is dependent on the reaction mixture pH.

PMID: 20709852 [\[PubMed - indexed for MEDLINE\]](#) PMID: PMC2950481 [Free PMC Article](#) [\[View\]](#)

mycoMINE on PMID: 20709852_Abtract [\(View\)](#) [\[View\]](#)

| Content | Type | Start | End | Features |
|--------------------|----------|-------|-----|---|
| cellobiohydrolase | Enzyme | 103 | 120 | <ul style="list-style-type: none"> ■ enzyme_alias: cellobiohydrolase ■ BRENDA_SystematicName: oligoxyloglucan reducing-end cellobiohydrolase ■ BRENDA_EcNumber: 3.2.1.150 ■ abbreviation_alias: - ■ google_search: http://www.google.com/search?q=cellobiohydrolase ■ BRENDA_RecommendedName: oligoxyloglucan reducing-end-specific cellobiohydrolase ■ SwissProt_ID: - ■ BRENDA's page: http://www.brenda-enzymes.org/php/result_flat.php4?ecno=3.2.1.150 |
| Magnaporthe oryzae | Organism | 143 | 161 | <ul style="list-style-type: none"> ■ 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 |

Found Entity

Entity Type

Entity Location

NLP-Provided Additional Information

- ▶ Extracted biomedical entities through NLP pipelines

PMID: 20709852 [PubMed - indexed for MEDLINE] PMCID: PMC2950481 Free PMC Article

[mycoMINE on PMID: 20709852 Abstract \(View\)](#)



| Content | Type | Start | End | Features |
|--|--------|-------|------|--|
| In contrast, TrCel6A exhibited slightly increased hydrolysis at pH 4.5, and hydrolysis was severely inhibited at pH 9.0. | pH | 1847 | 1967 | <ul style="list-style-type: none"> pH_alias: In contrast, TrCel6A exhibited slightly increased hydrolysis at pH 4.5, and hydrolysis was severely inhibited at pH 9.0. |
| cellobiohydrolase | Enzyme | 89 | 106 | <ul style="list-style-type: none"> enzyme_alias: cellobiohydrolase BRENDA_SystematicName: 4-beta-D-glucan cellobiohydrolase BRENDA_ECNumber: 3.2.1.91 abbreviation_alias: - google_search: http://www.google.com/search?q=cellobiohydrolase BRENDA_RecommendedName: cellulose 1,4-beta-cellobiosidase SwissProt_ID: O68438 BRENDA's page: http://www.brenda-enzymes.org/php/result_flat.php4?ecno=3.2.1.91 |

Semantic Entity Retrieval

Semantic Queries

- e.g., *“Which papers in this wiki mention an enzyme entity in their text?”*

```
{{#ask: [[hasType::Enzyme]]
| ?Enzyme = Enzyme Entities Found
| format = table
| headers = plain
| default = No pages found!
| mainlabel = Page Name
}}
```

| property | discussion | edit | history | delete | move | protect | watch | refresh |
|---|---|------|---------|--------|------|---------|-------|---------|
| Property:Enzyme | | | | | | | | |
|  Page Name |  Enzyme Entities Found | | | | | | | |
| DOIID: 10.1016/j.enzmictec.2006.03.017 Abstract | chitin deacetylase deacetylase polysaccharide deacetylase deacetylases | | | | | | | |
| DOIID: 10.1016/j.procbio.2007.01.007 Abstract | esterases FAEs xylanases | | | | | | | |

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design


ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

Semantic Navigation of NLP Results



navigation

- [Main page](#)
- [Community portal](#)
- [Current events](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)

search

toolbox

- [What links here](#)
- [Related changes](#)
- [Special pages](#)
- [Printable version](#)
- [Permanent link](#)
- [Semantic Assistants](#)
- [Browse properties](#)

[property](#)
[discussion](#)
[edit](#)
[history](#)
[delete](#)
[move](#)
[protect](#)
[watch](#)
[refresh](#)

Property:Enzyme

| Page Name | Enzyme Entities Found |
|-----------------|--|
| PubMed:19912637 | mannan endo-1,4-beta-mannosidase BK01 endo-1,4-beta-mannosidases mannanases 1,4-beta-D-mannan mannanohydrolase beta-mannanase hydrolase mannanase |
| PubMed:21861877 | β-glucosidase β-galactosidase cellobiohydrolases Cel7A/Cel6A endoglucanases Cel7B/Cel5A Cel7A Cel6A cellulase |

Pages using the property "Enzyme"

Showing 2 pages using this property.

P

[PubMed:19912637](#) + ⓘ mannan endo-1,4-beta-mannosidase, BK01, endo-1,4-beta-mannosidases, ...

[PubMed:21861877](#) + ⓘ β-glucosidase, β-galactosidase, cellobiohydrolases, ...

User Study

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

- ▶ Is the integration of text mining assistants in a wiki environment actually effective?
- ▶ User study within the Genozymes project context (www.fungalgenomics.ca)
 - ▶ **Goal:** Identifying and characterizing fungal enzymes
 - ▶ **Dataset:** 30 documents
 - ▶ **Users:** 2 expert biocurators
 - ▶ **NLP Service:** mycoMINE
 - ▶ **Measure:** Time spent on curation
 - ▶ **Method:** Comparison against time spent on manual curation

Average Curation Time

- ▶ Results:

| Abstract Selection | | Full Paper Curation | |
|--------------------|----------------|---------------------|----------------|
| no support | IntelliGenWiki | no support | IntelliGenWiki |
| 1 min. | 0.3 min. | 37.5 min. | 30.6 min. |

- ▶ **Conclusion:** IntelliGenWiki was indeed efficient and reduced the paper selection and curation time by almost **70%** and **20%**, respectively.

Introduction

The Wiki-NLP Integration

Hands-On

Application Examples

Conclusions

References

Conclusions

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

Add NLP capabilities to your wiki for a variety of use cases

- ▶ Allows use of existing text mining techniques in your wiki
- ▶ Find scenarios in which NLP assistance can be useful, e.g., Summarization
- ▶ Develop the actual NLP pipelines based on GATE
- ▶ Deploy the pipelines on a Semantic Assistants server
- ▶ Alternatively, use the existing text mining services in our public server
- ▶ Create machine-accessible information
- ▶ Add another party to the wiki user community: AI

Download and deploy the Wiki-NLP integration

- ▶ Deploy the Wiki-NLP servlet on a container, e.g., Tomcat or Jetty
- ▶ Install the Semantic Assistants MediaWiki extension on your wiki
- ▶ Configure the extension to point to the servlet endpoint

References



Phillip A. Laplante.

Requirements Engineering for Software and Systems.

Auerbach Publications, 1st edition, 2009.



Luisa Mich, Mariangela Franch, and Pierluigi Inverardi.

Market research for requirements analysis using linguistic tools.

Requirements Engineering, 9:40–56, 2004.

<http://dx.doi.org/10.1007/s00766-003-0179-8>.



Bahar Sateli, Elian Angius, and René Witte.

The ReqWiki Approach for Collaborative Software Requirements Engineering with Integrated Text Analysis Support.

In *The 37th Annual International Computer Software & Applications Conference (COMPSAC 2013)*, page 405–414, Kyoto, Japan, July 22–26 2013. IEEE.

<http://www.semanticsoftware.info/system/files/reqwiki-compsac13.pdf>.



Bahar Sateli, Marie-Jean Meurs, Greg Butler, Justin Powlowski, Adrian Tsang, and René Witte.

IntelliGenWiki: An Intelligent Semantic Wiki for Life Sciences.

In *NETTAB 2012*, volume 18 (Supplement B), pages 50–52, Como, Italy, 11/2012 2012. EMBnet.journal, EMBnet.journal.

References II



Bahar Sateli and René Witte.

Natural Language Processing for MediaWiki: The Semantic Assistants Approach.

In The 8th International Symposium on Wikis and Open Collaboration (WikiSym 2012), Linz, Austria, 08/2012 2012. ACM.



Axel van Lamsweerde.

Requirements Engineering: From System Goals to UML Models to Software Specifications.

John Wiley & Sons, 2009.



René Witte and Thomas Gitzinger.

Connecting Wikis and Natural Language Processing Systems.

In WikiSym '07: Proceedings of the 2007 International Symposium on Wikis, pages 165–176, New York, NY, USA, 2007. ACM.



René Witte and Thomas Gitzinger.

Semantic Assistants – User-Centric Natural Language Processing Services for Desktop Clients.

In 3rd Asian Semantic Web Conference (ASWC 2008), volume 5367 of LNCS, pages 360–374, Bangkok, Thailand, Feb. 2–5, 2009 2008. Springer.

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP
Integration

Requirements Analysis

System Design
Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References

References III

Introduction

Natural Language
Processing

Motivation

Semantic Assistants

The Wiki-NLP

Integration

Requirements Analysis

System Design

Implementation

Hands-On

Configuration

Information Extraction with
ANNIE

Semantic Queries

Application Examples

ReqWiki

Motivation

ReqWiki Design

ReqWiki Evaluation

IntelliGenWiki

Conclusions

References



René Witte, Ralf Krestel, Thomas Kappler, and Peter C. Lockemann.
Converting a historical architecture encyclopedia into a semantic knowledge
base.
IEEE Intelligent Systems, 25(1):58–66, January/February 2010.