Adding Text Mining Workflows as Web Services to the BioCatalogue

Georgios Kontonatsios, Ioannis Korkontzelos, BalaKrishna Kolluru, Sophia Ananiadou
Text Mining (TM)

- Chemistry tools
- Associations of biomedical concepts
- Systems Biology PathText
- Scientific workflows

Text Mining Systems

Online repositories of web services

Building Semantic tools for the Life Sciences using Text Mining
In IBM Journal of Research and Development:
U-Compare: a modular NLP workflow construction and evaluation system.
Background: Why workflows?

- Individual components may be of little interest
- Others require pre-processing steps

Diagram:
- Sentence Splitter
  - Part-of-Speech (PoS) Tagger
  - gene/protein NER
  - chemical compound NER

Workflow:
- Sentence Splitter
  - Part-of-Speech (PoS) Tagger
  - gene/protein NER
  - chemical compound NER

Part-of-Speech (PoS) Tagger

Sentence Splitter

ge/protein NER workflow

c hemical compound NER workflow
A framework that encourages interoperability in text analysis components

An architecture that facilitates workflow construction through roles and hierarchies
U-Compare is a TM platform built on top of UIMA.

It enables interoperability by defining a common Type System.
Background: U-Compare

- U-Compare allows the comparison and evaluation of workflows.
- Critical processes for constructing optimal workflows
 Integrating Biocatalogue with U-Compare

- Expose our solutions to the bioinformatics community
- Alternative access mechanism to U-Compare workflows
- Workflows are:
  - independent of programming languages or software libraries.
  - loosely-coupled components that can be reused in other applications.
User requirements and design objectives

- Integrate programatically with the web services

Web API

- Well documented web services

Documentation tools

- Easy-to-use tools
- Stay away from markup languages

Human readable access mechanism

Visualization of results

developers

user

documentation
tools

end-users

Georgios N. Kontonatsios
Architecture of the web application framework

UIMA/U-Compare REST Framework

- Descriptor Repository
  - Get XML descriptor
- Repository of resources
  - Get workflow
  - Resolve software dependencies
- U-Compare library
- Template Repository
  - Get template
- Apache Simple Server
- Inline Annotations
- Post-processing
- Stand-off annotations

Standalone workflow
- Pear package
  - Package in pear format
  - Register to framework

Workflow execution and result representation

Get XML descriptor
Get workflow
Submit text for analysis
Return result

Georgios N. Kontonatsios
### Architecture of the web application framework

#### Biomedical NER

<table>
<thead>
<tr>
<th>Web Service</th>
<th>Annotations</th>
</tr>
</thead>
<tbody>
<tr>
<td>NeMINE</td>
<td>gene, protein</td>
</tr>
<tr>
<td>ABNER</td>
<td>protein, DNA, RNA, cell line, cell type</td>
</tr>
<tr>
<td>GENIA</td>
<td>protein, DNA, RNA, Cell line, cell type, token, chunk</td>
</tr>
<tr>
<td>DECA</td>
<td>protein, DNA, RNA, Cell line, cell type, species</td>
</tr>
<tr>
<td>Yeast Metabliner</td>
<td>metabolites</td>
</tr>
<tr>
<td>OrgHub</td>
<td>organisms, habitats</td>
</tr>
</tbody>
</table>

#### Chemical NER

<table>
<thead>
<tr>
<th>Web Service</th>
<th>Annotations</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSCAR</td>
<td>gene, protein</td>
</tr>
<tr>
<td>OSCAR</td>
<td>chemical compound reaction, chemical expression, chemical adjective, ontology term</td>
</tr>
</tbody>
</table>
Conclusion & future work

- Alternative access mechanism to U-Compare workflows
- U-Compare Biocatalogue
- Expose TM tools to the Bioinformatics community
Conclusion & future work

Web application framework

- U-Compare workflows into web services
- Mirrors U-Compare library
- Post-processing mechanism
- Access interfaces for developers, end-users
- Templates adjusted to registered services
- Tools for documenting the web services

Access interfaces for developers, end-users

Templates adjusted to registered services

Tools for documenting the web services

U-Compare workflows into web services

Mirrors U-Compare library

Post-processing mechanism
Integration of the two frameworks is not automated yet

Framework presents delays in case of an overloaded network

XML is one of the ways that we use to model a web service

**Embody framework into U-Compare**

---

**Conclusion & future work**

![Graph](image-url)
Questions