

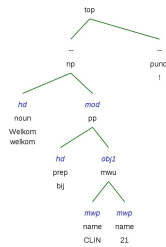
Syntactic Analysis of Dutch via Web Services

Erik Tjong Kim Sang
UNiversity of Groningen
erikt(at)xs4all.nl
11 February 2011

TTNWW project

Title: TST Tools voor het Nederlands als Webservices in een Workflow
Goal: develop easy access to Dutch NLP tools for researchers in the humanities and social sciences
Method: web services
Focus for Groningen: Alpino parser
Context: CLARIN project

Alpino: syntactic analysis of Dutch





Alpino Webdemo

ALPINO: AUTOMATIC SYNTACTIC ANALYSIS OF DUTCH

Sentences longer than 20 tokens (words) are ignored. The input is assumed to be a single sentence. Please type the sentence as you would do normally, with capitals at the beginning and for names etc.

SVG is used to display the resulting dependency structures. Recent browsers support this, but in some cases you might need to download an SVG plugin.

For more info on the Alpino Parser visit the [Alpino homepage](#)



Type in a sentence:
Welkom bij CLIN 21

Examples

<http://www.let.rug.nl/vannoord/bin/alpino>

What is a web service?

A web service is a variant of a remote procedure call that:

1. accepts a processing request, usually from remote machine
2. processes the request
3. returns some data

The data exchanged between the service and the client is usually formatted in XML

Existing web services for NLP

CLAM (Tilburg): create web service layer around existing NLP software

- <https://github.com/proycon/clam>

WebLicht (Germany): large-scale initiative to make NLP software available as web services

- <http://weblicht.sfs.uni-tuebingen.de/>

Data exchange protocols

A common data exchange protocol used for web services is SOAP

XML data is embedded in standard SOAP tags (also XML) before being sent from clients to server and vice versa

An alternative is to use REST (REpresentational State Transfer) which relies on existing web protocols: urls, http, etc.

Both CLAM and WebLicht use REST

Example of a SOAP message

```
<soap:Envelope xmlns:soap="http://www.w3.org/2003/05/soap...">
  <soap:Header>...</soap:Header>
  <soap:Body>
    <soap:Fault>
      <soap:Code><soap:Value>Sender</soap:Value></soap:Code>
      <soap:Reason><soap:Text>Message</soap:Text></soap:Reason>
    </soap:Fault>
    <sentence>Welkom bij CLIN 21!</sentence>
  </soap:Body>
</soap:Envelope>
```

Example of a TNS message in WebLicht

```
<?xml version="1.0" encoding="UTF-8"?>
<D-Spin xmlns="http://www.dspin.de/data" version="0.3">
  <tns:MetaData xmlns:tns="http://www.dspin.de/data/metadata">
    <tns:source></tns:source>
  </tns:MetaData>
  <tns:TextCorpus xmlns:tns="http://www.dspin.de/data/textcorpus"
    lang="nl">
    <tns:text>Welkom bij CLIN 21!</tns:text>
  </tns:TextCorpus>
  <!-- all other annotation -->
</D-Spin>
```

Extract from Alpino's XML output

```
<alpino_ds version="1.3">
  <node id="8" pos="adj" rel="mod" word="Frans"/>
  <node id="9" pos="noun" rel="hd" word="schrijver"/>
  ...
  <sentence>Alexandre Dumas père , Frans schrijver</sentence>
</alpino_ds>
```

Alpino is able to generate its output in XML format with syntactic information embedded in attributes of node tags

Authentication

We need some kind of authentication for the tools:

- some data access might require signing a license
- we like to keep track of who is using the tools and how often

SOAP protocols have been chosen for the authentication processes in the CLARIN projects

Demo

Tasks performed to create this demo:

- Fetch and start the Apache Tomcat web server
- Create Java interface (Java) between NLP tool and protocol (XML)
- Compile interface software to WAR file (Web application ARchive)
- Add WAR file to web server

Concluding remark

Alpino is not yet available as a web service but this project has only just started

Future Work

Develop web services for Alpino for TTNWW, when standards are available

Develop client software

THANK YOU FOR YOUR ATTENTION