A Web Service Interface for the Dutch Parser Alpino

Erik Tjong Kim Sang
University of Groningen

20 January 2012

TTNWW project (2010-2012)

Title: LST Tools for Dutch as Webservices in a Workflow

Goal: integration of existing Dutch language and speech technology tools

Framework: CLARIN: make resources and technology available for researchers in human and social sciences

Participants: Aletta, Antwerp, Ghent, Groningen, Huygens, INL, KADOC, KDC, Leuven, Meertens, Radboud, Tilburg, Twente and Utrecht

Groningen & TTNWW

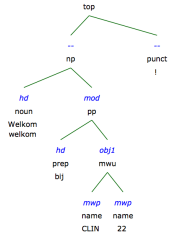
Focus for Groningen: Alpino parser

Method: build web service interface for Alpino

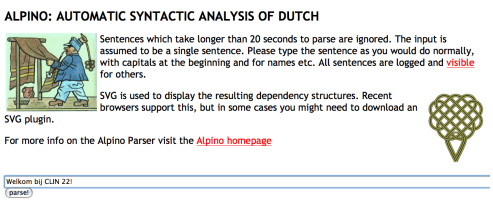
Note: this is the second of four TTNWW project talks in this session

The other three talks deal with ISOcat, FOLIA and infrastructures for resource integration

Alpino: syntactic analysis of Dutch



Alpino Webdemo



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

Steps taken in this project

- starting point: Alpino's server mode
- added server layer for processing more than one sentence in one run (Python script `alpino`)
- added server layer for tokenization and communication in XML format (Python script `server.py`)
- wrote command line text client (Python script `client.py`)
- working on XML input & output in Weblicht format (Python script `parse`)

Organization of the software

The three servers are running at Calligo, the computer cloud of BiG-Grid/SARA in Amsterdam: `https://www.cloud.sara.nl/`

The client software is presently only running at a machine in Groningen

Allowing another machine to access to the service requires a manual action from the admins at SARA, which takes several days

Demo

We present a demo of the current system:

- input via the text client
- input via the XML client

What needs to be done next?

1. expand and validate XML input & output
2. fix access for others: currently too limited
3. embed in larger framework (Weblicht)

And embed in other frameworks?

Concluding remarks

We are building a web service interface for Alpino

The access to the interface is limited and should be improved

The interface needs to be embedded in a larger framework

THE END