CLARIAH Language Resources

Jan Odijk

NoTaS 2017-10-12

Overview

- CLARIN-NL Portal, Services

- CLARIN-NL Portal, Data

- CLARIN Virtual Language Observatory
  - CLARIN Language Resource Switchboard

- CLARIAH MediaSuite

- CLARIAH DataLegend

- CLARIAH WP3

CLARIN-NL Portal, Services

- Search for tools, web services and applications via faceted search by tooltask, Research Discipline, language, and others

- Search, Syntax, Dutch

CLARIN-NL Portal, Data

- Search for data via faceted search by Research Discipline, annotations, language, and others

- Syntax, Syntactic annotation, Dutch

CLARIN Virtual Language Observatory (VLO)

- Search for data (and tools) via keyword and faceted search by multiple facets

- 1.6 million records

- e.g. "treebanks", language=Dutch -> 13 records

- Problems:
  - Amount of data make it difficult to find what you are looking for
  - Enormous differences in granularity
  - No good facets for tools (though some 400 tools are described)
    - UU is making a proposal for this
    - Together with software 'metadata' for some 70 (NL) tools

VLO: CLARIN Language Resource SwitchBoard

- Tools that can apply to data are suggested
  - Based on mimetype and language
    - From the metadata or guessed
    - E.g. this textfile, resource, and then e.g. Frog
- There is also a stand-alone version of CLRS

CLARIN Language Resource SwitchBoard

- Tools that can apply to data are suggested
  - Based on mimetype and language
    - From the metadata or guessed
    - E.g. this textfile, resource, and then e.g. Frog

CLARIAH Media Suite

- research environment in CLARIAH
- Targeted at media scholars (and other scholars who use audiovisual media)
- provides access to audiovisual collections and their contextual data.

[CLARIAH DataLegend](#) (Social Economic History)

- Link your datasets to other datasets across the world

- Supported by a range of tools:
  - Convert data to RDF: [Qber](#), [COW](#)
  - Reuse: [Vocabs](#) [LSD Dimensions](#)
  - Quality: BRWSR [Inspector](#)
  - Query: [Yasgui Grlc](#)
  - Store: [Druid](#)

WP3 Tools Updates (no portal yet)

- [PaQu](#)
- [PaQu SPOD](#) (under development)
- [GrETEL](#) (Utrecht variant. under development)
- OpenSoNaR+ (under development, soon at INT)
- …

ANANSI

- Instantiation of [Timbuctoo](#) repository system ([movie](#))

- Will contain objects ('what'), persons ('who') and locations ('where')

- As Linked Open Data

- CLARIN CMDI metadata converter to RDF (CMDI2RDF) can become part of Anansi

- Apparently, some language resources are already in there
  - Open Dutch Wordnet (VU)
  - DIAMANT (INT)