Text/Multimedia Processing & Analysis

CLARIAH Services (1)

1. **Natural Language Processing**
   - **User story** *As a scholar*, I want to automatically enrich my texts with linguistic annotation *in order to* facilitate further processing/analysis/visualisation steps for my research.
   - Covers a wide variety of NLP tasks & tools (Frog, Alpino, ucto, UDpipe-frysk)
   - Covers multiple languages (focus often on dutch), also consideration for historical language
   - **Proposed discussion points:**
     - Investing in state-of-the-art techniques (e.g. transformers)
     - Standardize web APIs for NLP tools?

CLARIAH Services (2)

1. Natural Language Processing
   - 1.1 **Spelling correction/normalisation**
     - OCR/HTR post-correction
     - **User story**: As a scholar, I want to correct or normalize text in order to make it more suitable (less errors/noise) for further-processing, whatever that may be; e.g. indexing for search, NLP processing, etc..
     - **Proposed discussion points:** PICCL maintenance issues
   - 1.2 **NLP Suites**
     - **Proposed discussion points:**
     - Suites vs individual components
     - Concerns of modularity, reusability and maintainability
   - 1.3 **Grapheme to Phoneme conversion**
2. (Annotated) Text Conversion
   - **User story**: As a scholar, I want to convert my (annotated) text document from one format to another *in order to* use my data with a tool that requires a different format, or because I want to store and archive it in a different format

CLARIAH Services (3)

4. Speech Recognition
   - **User story**: As a scholar, I want to create a speech transcripts for an audiofile so that I have a textual representation of it for browsing/close reading
   - **Proposed discussion points:**
   - Suitability of DANE in a wider context (text)?
   - User-story 2 (search)
   - Most dane worker links give a 404
5. Computer vision
   - **User story**: As a scholar, I want to use computer vision to explore data collections based on image features such as objects that are visible, shots or colours.

CLARIAH Services (4)

6. Linguistic Diagnostics Database
   - **User story**: As a scholar, I want to quickly find all arguments from the literature for or against a linguistic property of a word or construction
   - **Proposed discussion points:**
     - Interesting but very specific use case; less relevant from a cross-WP and overall infrastructure perspective
     - Suitability of PICCL for this is questionable due to sustainability issues
     - PDF processing components
7. Glossing service
   - **User story**: As a scholar, I want to use dutch examples in an English article and the gloss and translation should be added automatically
   - **Proposed discussion points:**
     - Move to NLP?
     - Suitability of PICCL for this is questionable due to sustainability issues