YEDDA: A Lightweight Collaborative Text Span Annotation Tool
Jie Yang, Yue Zhang, Linwei Li, Xingxuan Li
Singapore University of Technology and Design

Overview

Text Span Annotation: choose the span from input text and assign a tag on the chosen text span, e.g. NER/Chunking.

- **Example:**
  - Annotation: Person
  - Input text: Albert Einstein was born in Germany.

Manual Annotation Process:
1. Admin assigns the same data to multiple annotators.
2. Annotators label the data.
3. Admin analyzes the annotating results and gives feedbacks to the annotators.
4. Repeat previous steps until the inter-agreement is acceptable.
5. Start large-scale data annotation.

YEDDA:
- Designed for collaborative and efficient text annotation with GUI interface.
- Main advantages:
  - **Light:** size of core code ~80kB.
  - **Convenient:** lightweight; intuitive interface; Python 2.7; compatible with all main stream OS; no dependent package.
  - **Efficient:** it can significantly accelerate the annotation process.
  - **Comprehensive:** integrates various useful toolkits to accelerate annotating process for annotators and simplify the result analysis process for the administrator.

Annotator Interface:
- In Figure 2 which includes four main areas.
  - **Content Area:** up-left area. It shows the content of the input file, spans are colored based on different properties.
  - **Command Area:** down-left area, where is the entry of command line.
  - **Functional and Status Area:** in the right of content area, it contains 6 buttons which control the annotation settings. The cursors position and recommendation status are list below the buttons.
  - **Shortcut Mapping Area:** in the most right area. It presents the shortcut keys which are mapping to the specific span labels. This map can be customized with the functional buttons.

Annotator Client Functions:
- **Shortcut Annotation:** select text span and press shortcut key for annotating.
  - **Command Line Annotation:** it can parse the input command and annotate multiple spans at once.
  - **System Recommendation:** analyze annotated text to predict the potential spans in unannotated text to present predicted spans in green color to annotators confirm or ignore.

Speed Comparison:
- **YEDDA Speed:** our system can cut the annotation time in half compared with other SOTA annotation systems.
  - **YEDDA + Recommendation:** the system recommendation can give further 16.47% time reduction when annotating 100 sentences.

Administrator Interface:
- Figure 4: Administrator Interface.

Administrator Client Functions:
- **Multi-annotator analysis:** calculates the inter-agreement matrix for multiple annotators, as shown in Figure 5.
- **Comparison Report:** YEDDA can generate a detailed comparison report for each annotator pair. The disagreement annotated text is colored for comparison.

Figure 1: Collaborative annotation process.
Figure 2: Annotator Interface: Blue=Annotated entity; Green=Recommend entity; Orange=Current selected text.
Figure 3: Annotation speed comparison.
Figure 4: Administrator Interface.
Figure 5: Multi-annotator analysis result.
Figure 6: Multi-annotator analysis result.