



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 Location
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.

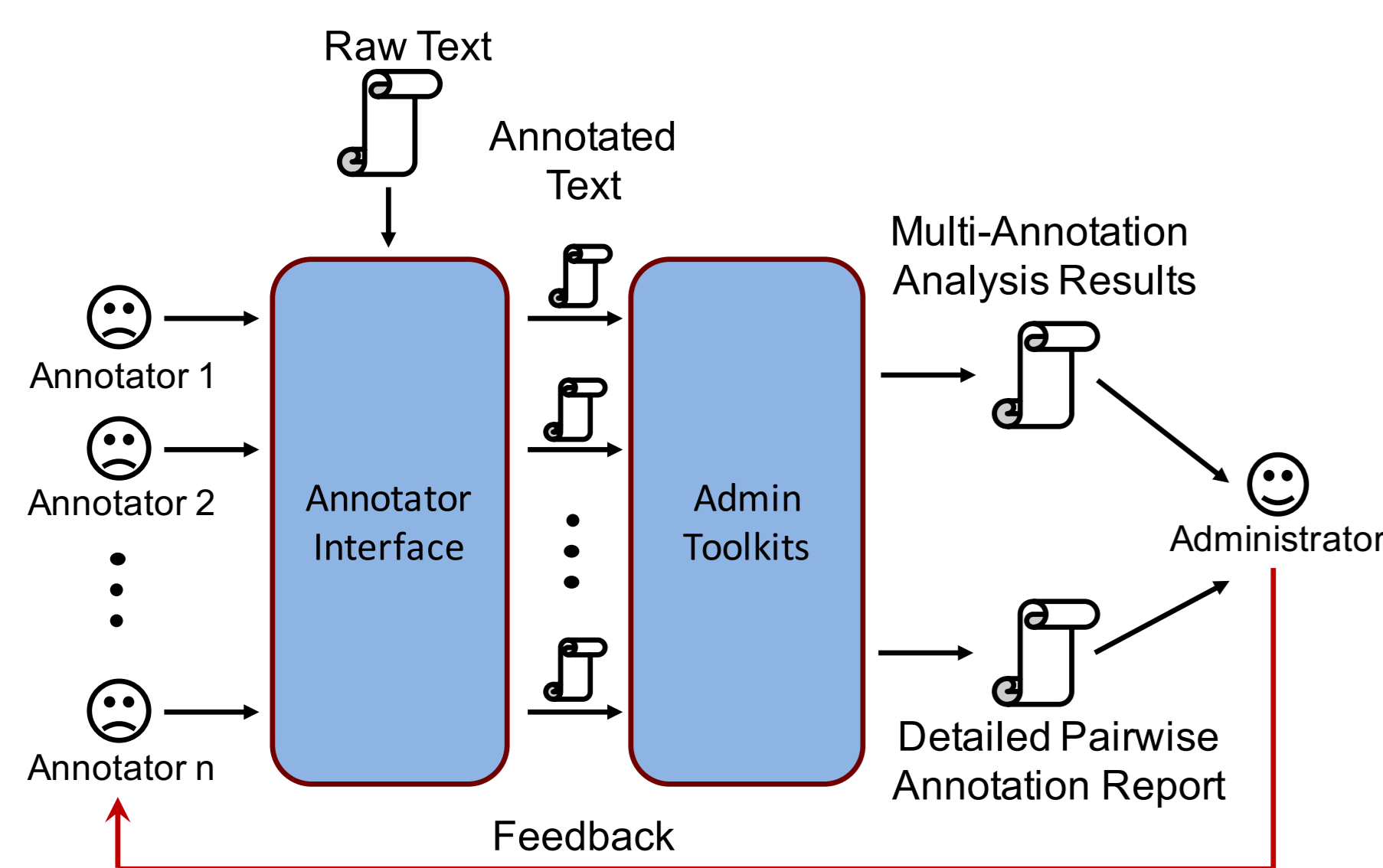


Figure 1: Collaborative annotation process.

YEDDA: is designed for collaborative and efficient text annotation with GUI interface. The main advantages are:

- **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:** it integrates various useful toolkits to accelerate annotating process for annotators and simplify the result analysis process for the administrator.

Annotator

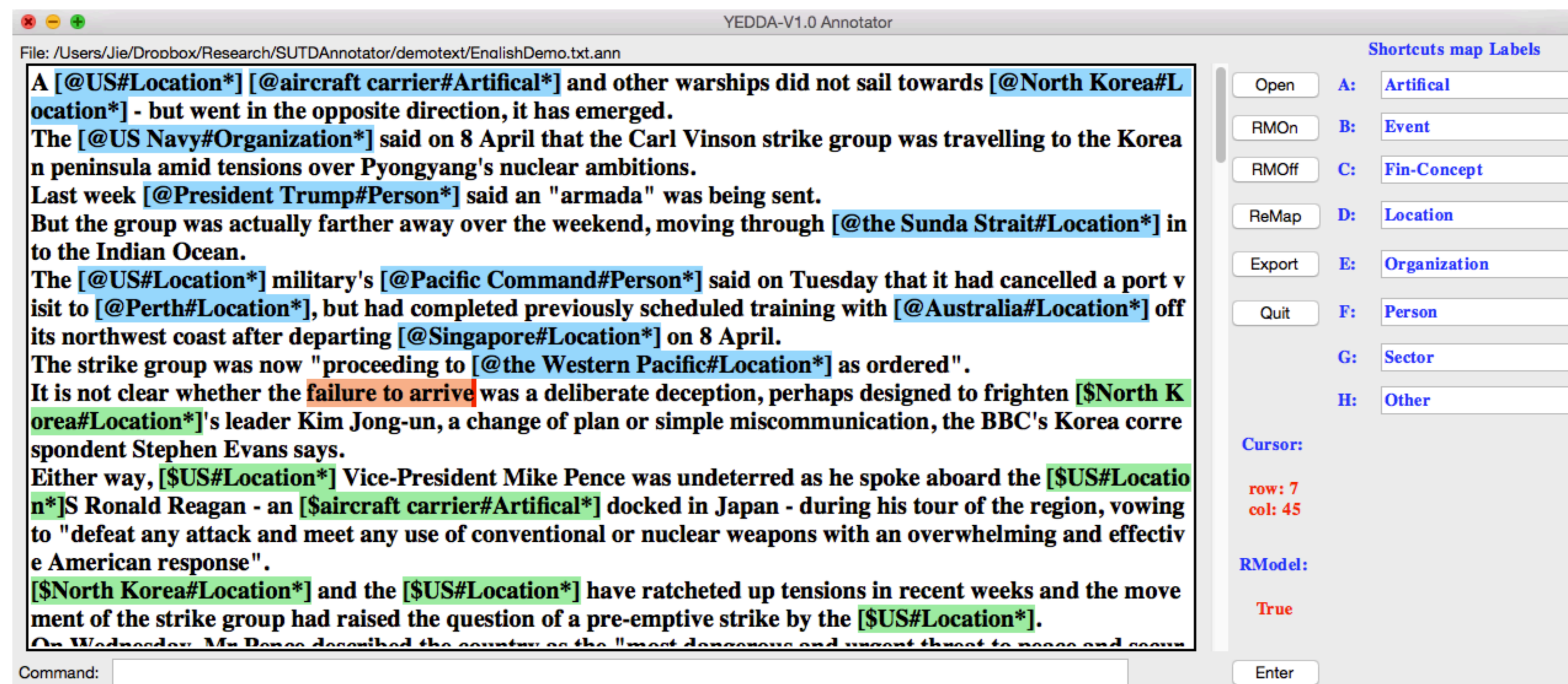


Figure 2: Annotator Interface: Blue=Annotated entity; Green=Recommend entity; Orange=Current selected text.

Annotator Interface: is shown 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 → predict the potential spans in unannotated text → present predicted spans in green color → annotators confirm or ignore.

Speed Comparison: (Figure 3)

- **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.

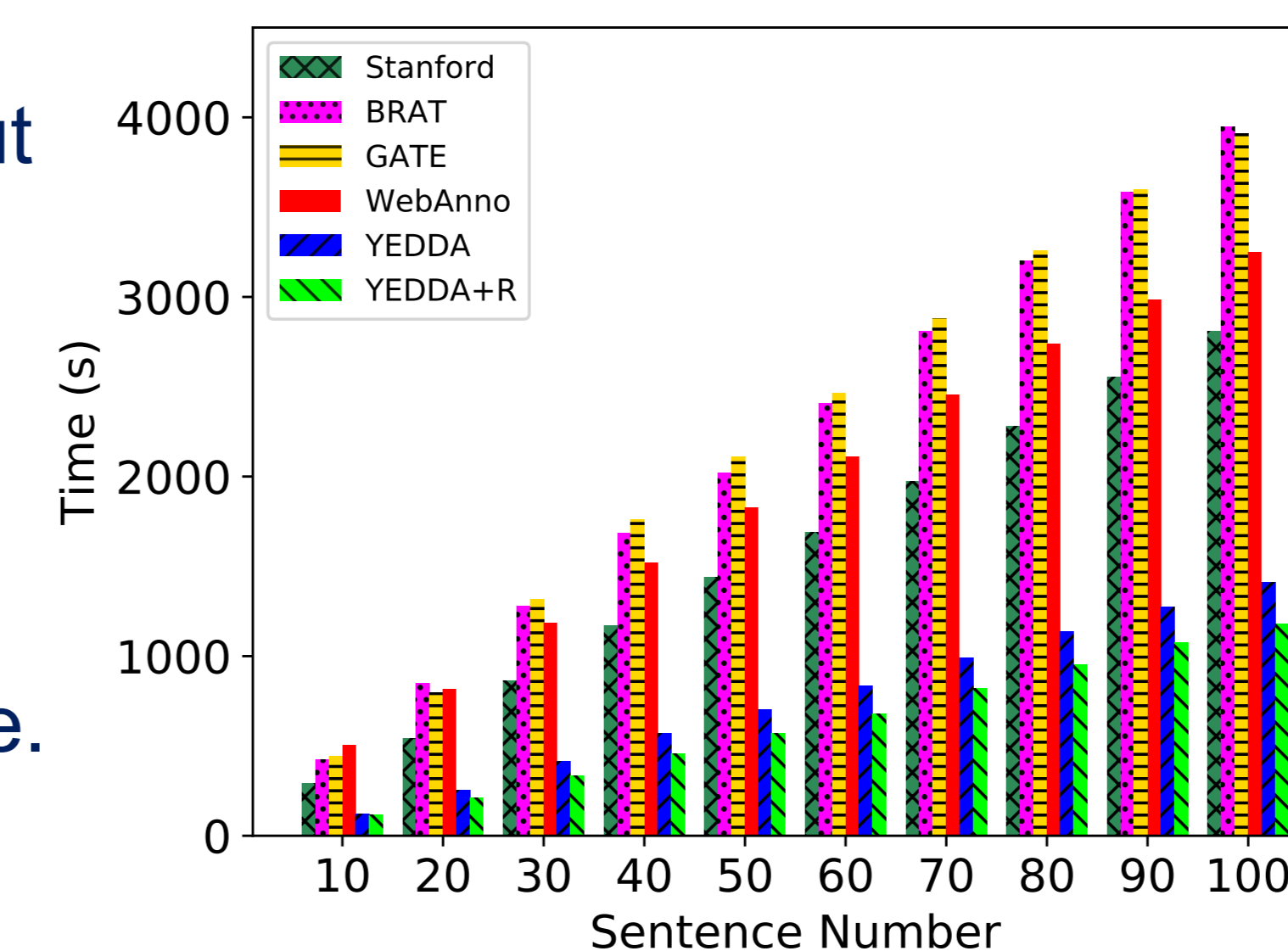


Figure 3: Annotation speed comparison.

Administrator

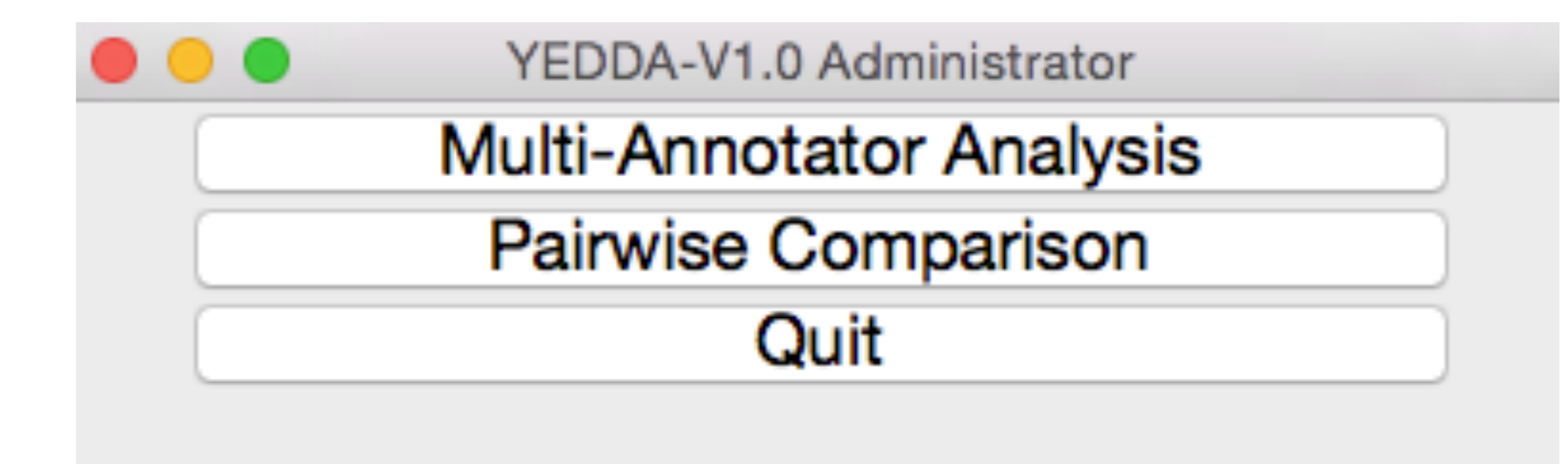


Figure 4: Administrator Interface.

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.

F:Entity/Chunk	UserA	UserB	UserC	UserD	UserE
UserA	100/100	64.1/74.5	70.5/80.9	52.6/62.5	59.0/67.3
UserB	64.1/74.5	100/100	64.8/74.9	58.8/67.1	56.1/66.7
UserC	70.5/80.9	64.8/74.9	100/100	54.7/64.6	57.9/69.2
UserD	52.6/62.5	58.8/67.1	54.7/64.6	100/100	63.7/75.9
UserE	59.0/67.3	56.1/66.7	57.9/69.2	63.7/75.9	100/100

Figure 5: Multi-annotator analysis result.

Annotation Comparison Report
SUTDNLP Group
Singapore University of Technology and Design

1 Overall Statistics

File1 color: Blue; Dir: ~/demotext/EnglishUserA.txt.ann
File2 color: Red; Dir: ~/demotext/EnglishUserB.txt.ann

Table 1. Statistics for two annotations, assume File1 as gold standard

P/R/F (%)	Entity	Boundary
Artificial	74.36/50.0/59.79	-
Event	Nan/0.0/Nan	-
Fin-Concept	57.36/68.07/62.26	-
Location	50.0/100.0/66.67	-
Organization	66.67/82.93/73.91	-
Other	57.14/44.44/50.0	-
Person	91.3/67.74/77.78	-
Sector	56.52/76.47/65.0	-
Overall	61.96/66.36/64.08	72.05/77.16/74.52

2 Detail Content Comparison

Blue: only annotated in File1.
Red: only annotated in File2.
Green: annotated in both files.

But the group was actually farther away over the weekend, moving through the Sunda Strait into the Indian Ocean. The US military's Pacific Command said on Tuesday that it had cancelled a port visit to Perth, but had completed previously scheduled training with Australia off its northwest coast after departing Singapore on 8 April. The strike group was now "proceeding to the Western Pacific as ordered". It is not clear whether the failure to arrive was a deliberate deception, perhaps designed to frighten North Korea's leader Kim Jong-un, a change of plan

Figure 6: Multi-annotator analysis result.