Annotated Corpus Development

Challenges of annotation tasks
- Highly time-consuming
- Requires qualified expert judgment
- Prone to inter-annotator variability

Study Questions:
- Do pre-annotations help accomplish an annotation task? Impact on annotation time, ease of work…
- Do pre-annotations influence annotators in a negative way? Impact on annotation quality, bias…

Material and Methods

10,000 PubMed queries

Divide into 50 batches

Validation of annotation scheme by all 7 annotators

Annotation from scratch (1 annotator per batch)

Automatic pre-annotations (1 annotator per batch, except 1 common batch)

Annotation interface
Knowtator was used for performing annotations and computing Inter Annotator Agreement

Pre-annotations
Pre-annotations were obtained using MetaMap for the biological categories. PubMed query tags and rules were used for the abbreviation and bibliographic data categories.

Acknowledgments
This study was supported by the Intramural Research Programs of NIH/NLM.

Results

Figure 1. Results on batches with vs. without pre-annotations: visible benefits in terms of number of actions, annotation time, and consistency.

Figure 2. Evolution of annotation time as the task progresses for the seven annotators.

To quantify the influence of training vs. pre-annotations, mixing the batches’ distribution would be a helpful experimental setting.

Conclusions on pre-annotations
- Helpful
- Do not induce negative bias
- Welcomed by annotators

References
2. Islamaj Doğan et. al. Understanding PubMed user search behavior through log analysis. Database, 2009