Information Extraction of Seminars – Report

Entity Tagging

Tagging Times

Tagging Speakers

Tagging Locations

Tagging Sentences and Paragraphs

Evaluation  
All measures are given to 4 significant figures.

All 184 Text Files

Speaker Accuracy: 11.59%

Location Accuracy: 46.87%

Start Time Accuracy: 15.55%

End Time Accuracy: 60.87%

Sentence Accuracy:

Paragraph Accuracy:

Overall Accuracy: 13.13%

Overall Precision: 17.90%

Overall Recall: 30.22%

Overall F measure: 0.0%

First 92 (301-392) Text Files

Speaker Accuracy: 9.196%

Location Accuracy: 57.01%

Start Time Accuracy: 14.76%

End Time Accuracy: 45.65%

Sentence Accuracy:

Paragraph Accuracy:

Overall Accuracy: 12.56%

Overall Precision: 17.07%

Overall Recall: 28.44%

Overall F measure: 0.0%

Last 92 (393-484) Text Files

Speaker Accuracy: 13.98%

Location Accuracy: 36.72%

Start Time Accuracy: 16.33%

End Time Accuracy: 76.09%

Sentence Accuracy:

Paragraph Accuracy:

Overall Accuracy: 13.70%

Overall Precision: 18.72%

Overall Recall: 32.00%

Overall F measure: 0.0%

Ontology Construction

Future Improvements

I found myself spending more time working on entity tagging then ontology construction. If I were to re-do this assignment I’d ensure I split my time more equally between the two tasks.

I would only tag speakers that have the greatest number of occurrences in the text

I would tag locations outside of the ‘where:’ section of the email