|  |  |  |
| --- | --- | --- |
| **Task** | **Person** | **Date and Time** |
| Define project idea and scope | Dirk, Huda | 11/08 11:00 – 14:00 |
| Create training data (emails) | Dirk, Huda | 11/08 14:00 – 14:30 |
| Load training data in the code | Dirk | 11/08 14:30 – 15:00 |
| Define features | Huda | 11/08 14:30 – 15:00 |
| Create chunker | Dirk, Huda | 11/08 15:00 – 18:00 |
| Create features | Dirk, Huda | 11/10 14:00 – 16:00 |
| Finalize and comment code | Dirk, Huda | 11/10 16:00 – 17:00 |
| Analyze results | Dirk, Huda | 11/10 17:00 – 17:30 |
| Write up of results and report | Dirk, Huda | 11/10 17:30 – 18:00 |

We choose suggestion 1 of the Information Extraction assignment. Based on emails with invitations of events and conferences or talks we want to extract information about the location where the event will take place.

For example: There is an event on Friday in South Hall 202. The NE classifier should find ‘South Hall 202’.

Features that we think are important:

* Words have capitals except for stopwords
* There are numbers included
* Most words are nouns
* There are brackets ()
* Most words won’t be in the dictionary
* Most words are names NP
* The locations often include hall, school or house

We train the classifier by saving the email invitations in separate files and save them by the location name. The classifier is trained by the features mentioned above.