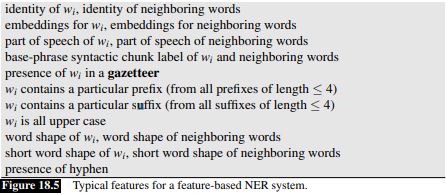
There has not been much previous work that has been done on this topic, but each of the individual aspects of the project have been researched in depth by other parties. Given that this is an information extraction task, this literature survey will give a background on different techniques that have been used successfully in that past. The information extraction task is broken down into named entity recognition and relationship extraction.

\section Named Entity Recognition

The usual approach to named entity recognition is supervised learning models which classify tokens in the text with a label. A token in most cases refers to a singular word. Most commonly these models use the BIO scheme, where they will label tokens as either the beginning, inside, or outside of a class. To make the decisions on what to classify a token as they model need additional information, this comes in the form of features. An example of features that could be used is given in Speech and Language Processing by Daniel Jurafsky & James H. Martin section 18.1.2, figure 18.5.



This is all additional information that helps to inform the models decision. When data is given to the model either for training or classification, it must be converted into this format. One thing to note is that this is a general set of features for general named entity recognition, as with this project we only want to identify either objects or sizes, we could tailor the features specifically to our need. For example, if choosing features to label tokens as a size, one helpful feature would be if that token included a unit of size. This would greatly increase the accuracy of the model (assuming that our gazetteer is large and accurate) for determining sizes but would be near useless when determining if a token is an object.