

## 1 Summary of Discussions

- Brush up on word sense disambiguation techniques currently in use.
- Apply WSD on the DNA level and use the information from the RNA and protein levels for it.
- Define and use ontological relations between the above mentioned layers to help in knowledge extraction.
- Use unsupervised methods to get the senses and then use supervised methods.

## 2 Papers Read

Roberto Navigli. 2009. Word sense disambiguation: A survey. *ACM Comput. Surv.* 41, 2, Article 10 (February 2009), 69 pages.

This paper gives a brief summary of the techniques that have been used to solve the Word sense disambiguation problem. It describes different methods for the selection of word senses as well as representation of contexts. The paper also differentiates between different types of classification methods:

- **Supervised:** Decision Lists, Decision Trees, Naive Bayes, Neural Networks, Instance-based, SVMs, Ensemble methods
- **Semi-supervised:** Bootstrapping, Active Learning
- **Unsupervised:** Context Clustering, Word Clustering, Co-occurrence graphs, Other graph based approaches

The paper also lists out other approaches and topics that are domain specific to languages.

## 3 New Directions

- Will need to define what a neighborhood means in this domain. In NLP, we generally assume that only nearby words affect the word under consideration, which might not be true in this domain.
- Will need to think about the baselines that we will use. Can use methods already implemented in this domain as well as naive methods from WSD.

## 4 Plan of Action

- Literature survey specific to the problem statement.
- Brush up on the biology related concepts.