

Assignment-Discussion Vector Based WSD

<name-1>, <roll no.-1>

<name-2>, <roll no.-2>

...

<date>

Problem Statement: Part-1: Vector based extended Lesk

- Given a sequence of words, produce the synset IDs of only NOUNS
- Data SemCor
- First Technique to be used: Extended Lesk
- 5-fold cross validation
- See if the baselines are crossed (WFS, MFS)

First Baseline

- Most Frequent Sense (MFS)
- <Give Values>

Second Baseline

- Wordnet First Sense (WFS)
- <Give Values>

Performance report of Extended Lesk

- Precision
- Recall
- F1-score
- On 5 folds

Part-2: Page Rank Based

- Repeat the slides as given in Extended Lesk for this approach
- However, PR is a all-words-wsd
- You have to give performance figures only for nouns

Confusion Cases (for EL)

- Present and discuss confusion cases
- For example, is “finance” sense of *bank* confused with “water body side” sense of *bank*?
- Give 5 such observations

Confusion Cases (for PR)

- Present and discuss confusion cases
- For example, is “finance” sense of *bank* confused with “water body side” sense of *bank*?
- Give 5 such observations

Interpretation of confusion (error analysis: EL)

- <try giving reasons>

Interpretation of confusion (error analysis: PR)

- <try giving reasons>

Data Processing Info

- Word Vectors and how you obtained them
- ONLY word2vec

MUST: GUI

- Give screenshot