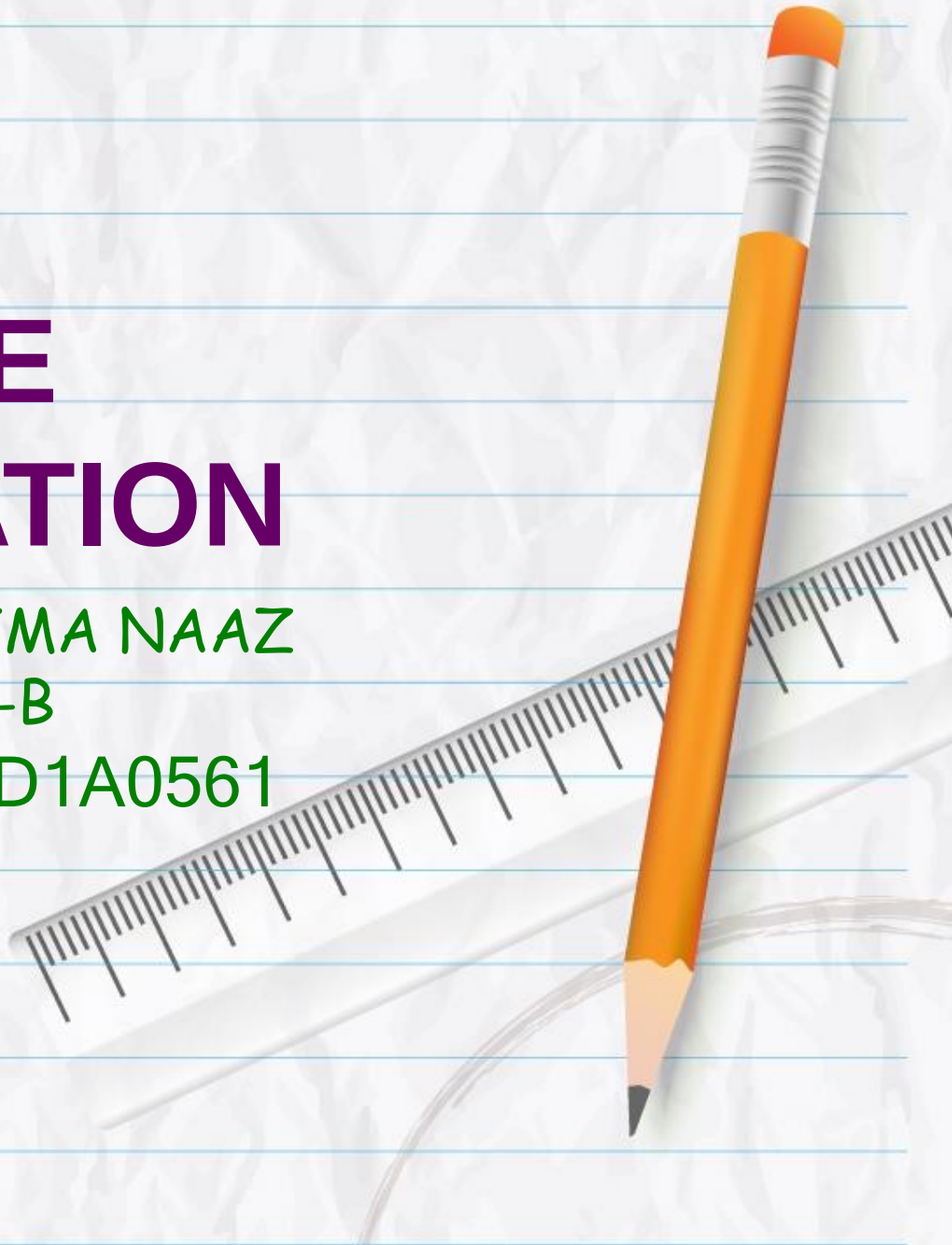
WORD SENSE DISAMBIGUATION

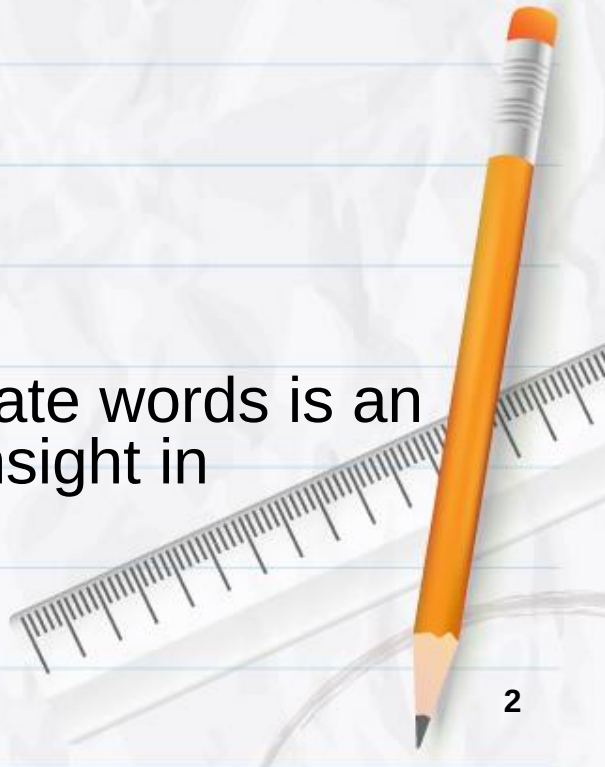
by NAJMA NAAZ
CSE-B
10BD1A0561



Motivation for Word Sense Disambiguation (WSD)

The computational identification of meaning for words in context

- Many tasks in natural language processing require disambiguation of ambiguous words.
  - Question Answering
  - Information Retrieval
  - Machine Translation
  - Text Mining
  - Phone Help Systems
- Understanding how people disambiguate words is an interesting problem that can provide insight in psycholinguistics.



Task Description

- Seletion of 'Word Senses'
- External Knowledge Sources
- Representation of Context
- Seletion of Automatic Classification Methods

Seletion of 'Word Senses'

- Sense Inventory
- Enumerative Approach
- Generative Approach
  - QUALIA
    - Formal
    - constitutive
    - telic
    - agentive

External Knowledge Sources

- Structured resources
  - Thesauri
  - Machine readable dictionaries
  - Ontologies
- Unstructured Resources
  - Corpora
  - Raw Corpora
  - Sense annotated corpora
  - Collocation Resources
  - Other Resources
- WordNet
- SemCor

Representation of Context

- Tokenization
- Parts of Speech Tagging
- Lemmatization
- Chunking
- Parsing

Seletion of Automatic Classification Methods

A. Supervised WSD - Approaches

1) Decision Lists
2) Decision Tree
3) Naive Bayes
4) Neural Networks
5) Exempler Based or Instance Based learning
6) Support Vector Machines
7) Ensemble Methods

Seletion of Automatic Classification Methods

B. UnSupervised WSD - Approaches

1) Context Clustering
2) Word Clustering
3) Co-Occurence Graphs

Seletion of Automatic Classification Methods

C. Knowledge Based WSD - Approaches

1) Overlap of Sense Definitions
2) Selectional Preferences
3) Structual Approaches

D. Other Approaches

1) Determining Word Sense Dominance
2) Domain Driven Disambiguation
3) WSD from Cross - Lingual Evidence

Evaluation Methodologies

- Evaluation Measures
  - *Coverage C*
  - *Precision P*
  - *Recall R*
- Baselines
  - Random Baselines
  - The First Sense Baseline
- Lower and Upper Bounds

Evaluation: The SenseEval/ SemEval Competitions

- Standardized international “competition” on WSD.
- Organized by the Association for Computational Linguistics (ACL) Special Interest Group on the Lexicon (SIGLEX).
- Three held, fourth planned:
  - Senseval 1: 1998
  - Senseval 2: 2001
  - Senseval 3: 2004
  - Senseval 4: 2007

Applications

- Information Retrieval
- Information Extraction
- Machine Translation
- Content Analysis
- Word Processing
- Lexicography
- The Semantic Web

Conclusion

- WSD is a hard task - dealing with full complexities of language
- Performance Tuning should be investigated further in the evaluation campaigns