NAME ENTITY RECOGNITION

- Introduction
- NER systems
- Main challenges

INTRODUCTION

Named-entity recognition (**NER**) consists of locating and classifying named entities in text into pre-defined categories such as:

- Names of persons
- Organizations
- Locations
- Expressions of times
- Quantities
- Monetary values
- Percentages

INTRODUCTION

Taking as input an unannotated block of text

Jim bought 300 shares of Acme Corp. in 2006.

And producing as output an annotated block of text that highlights the names of entities:

[*Jim*]$_{Person}$ *bought 300 shares of* [**Acme Corp.**]$_{Organization}$ *in* [**2006**]$_{Time}$.

INTRODUCTION

NER includes two tasks:

1. Detection of the name. Can be simplified to a segmentation problem, similar to chunking. Names are defined to be a contiguous set of tokens.

Bank of America is segmented as a name

2. Classification of the name by the type of entity they refer to: person, organization, location and other.

Certain hierarchies of named entity types have been proposed.

NER SYSTEMS

NER systems use several techniques

- Hand-crafted linguistic grammars.

    - Good precision

    - Low recall

    - High cost. Months of experienced computational

        linguistics

- Statistical models, i.e., Machine learning, Hidden Markov models

    -   Require a large amount of manually annotated training data

- Hybrid models

    - They avoid the need of manually annotated data

MAIN CHALLENGES

- NER systems developed for one domain do not typically perform well on other domains.
  - The same problem for both techniques: rules based and statistical models
- NER systems requires months of computational linguistics.
  - For writting grammars and annotating corpus
  - Supervised techniques seem a solution
- Scaling up to fine-grained entity types