Entity Recognition and Relation Extraction

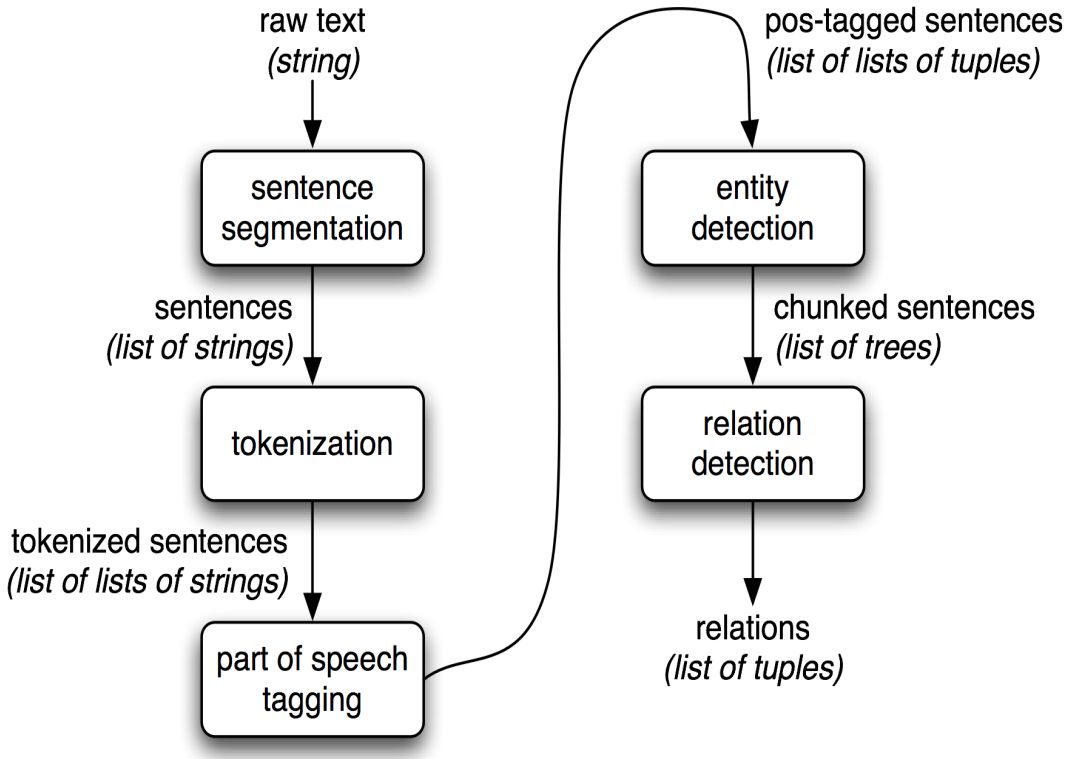
Sai Kiran Naragam
N100638

Under the Guidance of
Mr. Uday Kumar Ambati
Asst. Professor
Dept. of Comp. Sci. & Engg.
RGUKT Nuzvid

Presentation Outline

Aim

To recognise entities(Named Entites) in the given text and extracting relationships.



www.nltk.org/book/

Presentation Outline

Intro

In 1917, Einstein applied the general theory of relativity to model the large-scale structure of the universe. He was visiting the United States when Adolf Hitler came to power in 1933 and did not go back to Germany, where he had been a professor at the Berlin Academy of Sciences. He settled in the U.S., becoming an American citizen in 1940. On the eve of World War II, he endorsed a letter to President Franklin D. Roosevelt alerting him to the potential development of "extremely powerful bombs of a new type" and recommending that the U.S. begin similar research. This eventually led to what would become the Manhattan Project. Einstein supported defending the Allied forces, but largely denounced using the new discovery of nuclear fission as a weapon. Later, with the British philosopher Bertrand Russell, Einstein signed the Russell-Einstein Manifesto, which highlighted the danger of nuclear weapons. Einstein was affiliated with the Institute for Advanced Study in Princeton, New Jersey, until his death in 1955.

Tag colours:

LOCATION TIME PERSON ORGANIZATION MONEY PERCENT DATE

Image: *http://www.europeana-newspapers.eu/named-entity-recognition-for-digitised-newspapers/*

Entity Recognition

We can achieve this in two ways:

Entity Recognition

We can achieve this in two ways:

- `Procedure 1`
  - PoS Taggins
  - Named Entity Chunking

Entity Recognition

We can achieve this in two ways:

- `Procedure 1`
  - PoS Taggins
  - Named Entity Chunking
- `Procedure 2`
  - Named Entity Extraction as Tagging

Intro. to PoS Tagging

Part-of-Speech Tagging

INPUT:

Profits soared at Boeing Co., easily topping forecasts on Wall Street, as their CEO Alan Mulally announced first quarter results.

OUTPUT:

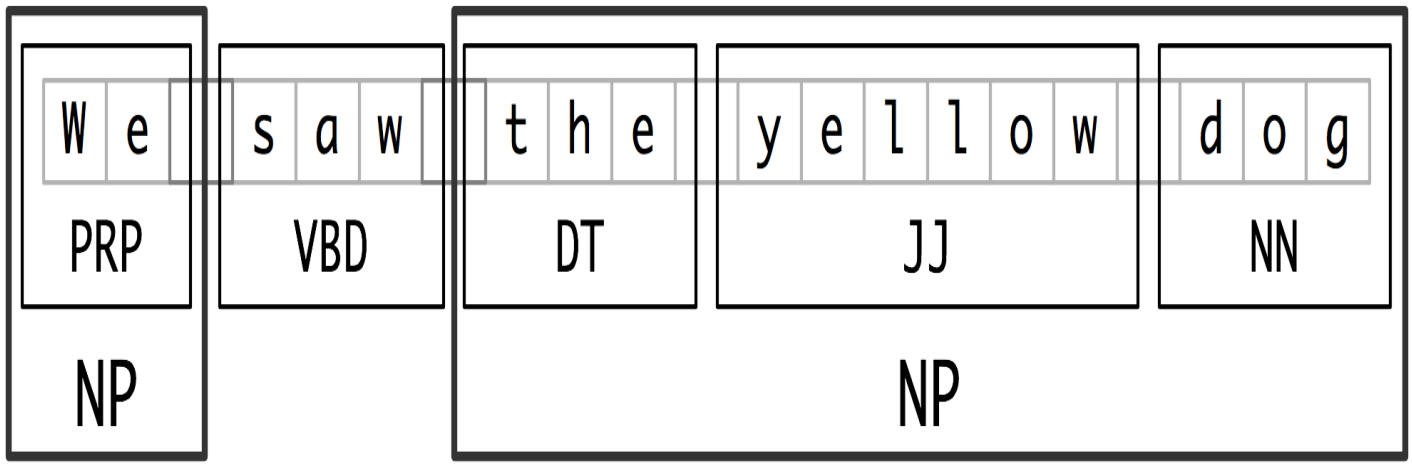
Profits/N soared/V at/P Boeing/N Co./N ,/, easily/ADV topping/V forecasts/N on/P Wall/N Street/N ,/, as/P their/POSS CEO/N Alan/N Mulally/N announced/V first/ADJ quarter/N results/N ./.

N = Noun
V = Verb
P = Preposition
Adv = Adverb
Adj = Adjective
...

PoS Tagging Contd..

Here we used NLTK's recommended PoS Tagger.It uses the Penn Treebank tagset.

Intro to Named Entity Chunking



www.nltk.org/book/

Named Entity Chunking Contd...

- We can use *hand-written rules(* **Ex: regular expressions** to chunk tags in to a named-entity. Or we use a trained chunker.
- In the case we used NLTK's recommended Chunker. Which is trained on ACE(Automatic Context Extraction).

Entity Extraction itself as Tagging

Example: Bikel et. al 1999 (Named Entity Recognition)

Named Entity Extraction as Tagging

INPUT:

Profits soared at Boeing Co., easily topping forecasts on Wall Street, as their CEO Alan Mulally announced first quarter results.

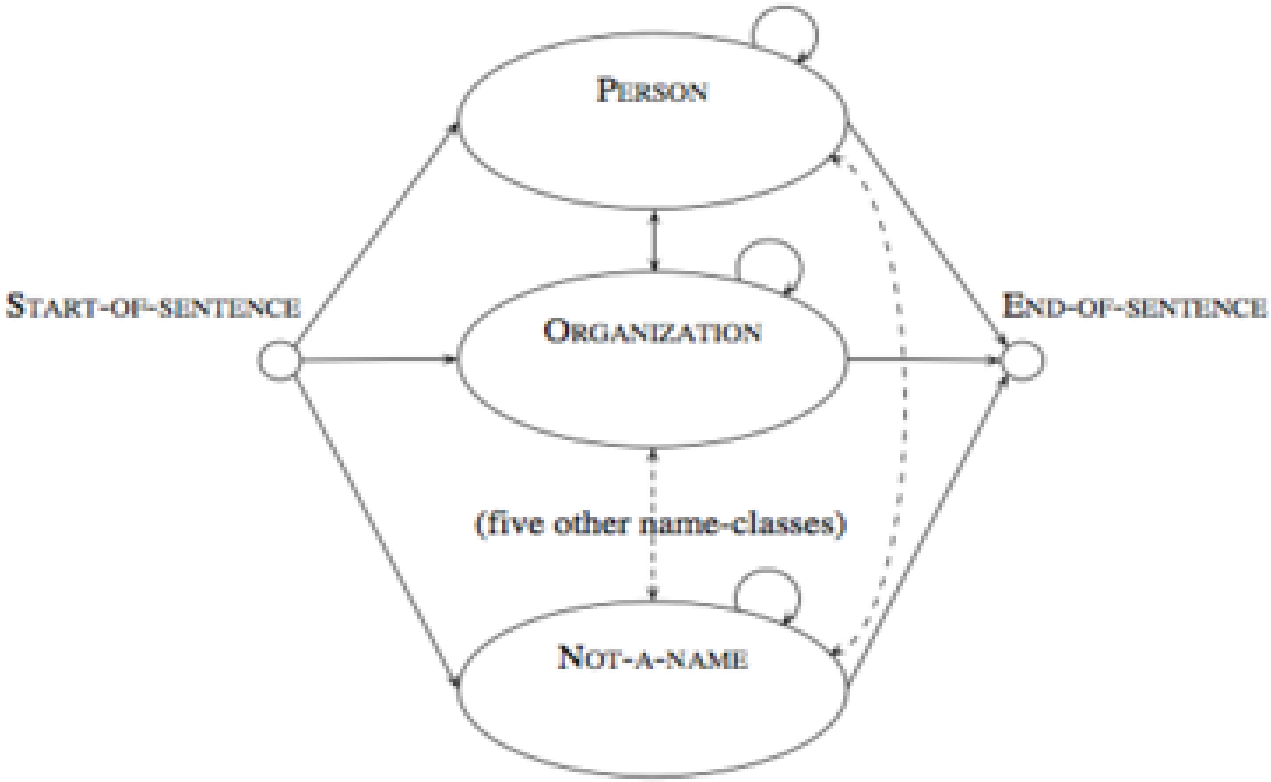
OUTPUT:

Profits/NA soared/NA at/NA Boeing/SC Co./CC ,/NA easily/NA topping/NA forecasts/NA on/NA Wall/SL Street/CL ,/NA as/NA their/NA CEO/NA Alan/SP Mulally/CP announced/NA first/NA quarter/NA results/NA ./NA

NA = No entity
SC = Start Company
CC = Continue Company
SL = Start Location
CL = Continue Location
...

Entity Extraction as Tagging Contd..

Bikel et al. used Hidden Markov Models to tag.



http://curtis.ml.cmu.edu/w/courses/index.php/

Presentation Outline

Intro to Relation Extraction

What kind of relationship is there between entities

PERSON

Founder?

Investor?

Member?

Employee?

President?

ORGANIZATION

Intro to Relation Extraction

- Here we are using *hand-written patterns(* **regular expressions** *)* to extract the relationship.
- We may use the so-called *learning technique*. But we may need huge hand-labeled training data(in case of supervised) or very advanced techniques(in case of unsupervised).

How ?

- Form relation-tuples of the form *(subj, filler, obj)*
- *subj,obj* are Entities and *filler* is text.
- Apply rules on *filler* to extract relation.

Applications of Relation Extraction

Concentrated application is **Question Answering**
where are different facilities located
who is employed by what company.

Presentation Outline

Summary of Work

- Technologies: Python,NLTK.
- Tested: BBC business text.
- Worked only with few class PERSON,LOCATION,ORGANIZATION,GPE.
- Looking forward to work with 7 classes.
- Started workin with **Stanford Named Entity Recognizer**.
- Theoritical study of HMMTagger,Maximum Entropy Markov Model Tagger(*not completed*) and Perceptron Tagger(*not completed*).
- Generative vs. discriminative learning models.

Presentation Outline

Future Implementations

- To do everything done so far for Indian languages.
- Tried but left the work due to some difficulties observed.

Thank You