

# CS5560 Knowledge Discovery and Management

## In-Class-Exercise (ICE-1B)

June 5, 2017

Name: SAI TEJA MAKANI

ClassID: 12

1. Given the text example below, show the following NLP operations.
  - a. Sentence detection/Tokenization
  - b. Lemmatization
  - c. POS tagging/Chunking
  - d. Parsing
  - e. Name Entity Recognition
  - f. Co-reference Resolution
2. Summarize and draw a knowledge graph.

CHICAGO (AP) — Citing high fuel prices, United Airlines said Friday it has increased fares by \$6 per round trip on flights to some cities also served by lower-cost carriers. American Airlines, a unit AMR, immediately matched the move, spokesman Tim Wagner said. United, a unit of UAL, said the increase took effect Thursday night and applies to most routes where it competes against discount carriers, such as Chicago to Dallas and Atlanta and Denver to San Francisco, Los Angeles and New York.

June 12, 17

# C55560 Knowledge Discovery and Management

## In-class-Exercise (ICE-IB)

SAI TEJA MAKANI

Number - 12

1.

### a. Sentence detection / Tokenization:

→ A text can be Tokenized into different sentences by "." ~~and~~ So in the give text we have 3 Sentences only.

For non-English it is incredibly difficult.

### b. Lemmaization: In linguistics is the process of grouping the inflected forms of a word so they can be analysed as a single item, identified by the word's lemma.

citing	—	cit
increased	—	increase
competes	—	compete
carriers	—	carrier
applies	—	appli

Sometime Lemmaization results in a non-meaningful words too.

### c. POS tagging / chunking: Finding Parts of speech is

on the main challenge in NLP. An example is shown below.

NNP	VBD	VBD	IN	NNP	TO	NNP	CC	NNP	NNP
John	was	born	in	Liverpool,	to	Julia	and	Alfred	Lennon



VBD VBD NNP VBD NNP NNP VBD DAY  
 citing high fuel prices, United Airlines said Friday  
 ITO VBD VBD VBD TO NUM VBD  
 it has increased fares by \$6 per round trip on  
 NNP TO HUBD VBD VBD VBD  
 flights to some cities also served by lower-cost  
 VBD  
 carriers.

d. Passing: It is the process of solving trees using  
 features like Lexicals and Structure. predicting  
 over search and greedy, Spanning trees,

e. Name Entity Recognition: Identifying names, entities  
 in a text Based on context of the text.

Facility — United Airlines  
 — American Airlines  
 — AMR  
 — United  
 — UAL

Location — Chicago  
 — Dallas  
 — Atlanta  
 — Denver  
 — San Francisco  
 — Los Angeles  
 — New York

Person — Tim Wagner  
 Product — Fuel  
 — Flights



F. Co-reference Resolution: finding the citations that are referring from the same ~~to~~ different sentences to the same Named Entity. co-ref

ex: John read a book, He wrote summary of it.  
co-ref

2. Knowledge graph:

