CS5560 Knowledge Discovery and Management

In-Class-Exercise (ICE-1B)

June 5, 2017

Name: Sree Lakshmi Nandaramudi

ClassID: 17

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.

Scanned by CamScanner

1. NLP operations :

a) Sentence detection / Tokenization :

The sentence detection is used to detect the sentences based on the punctuation character. Let us divide the sentence based on the period. The sample text is segmented into sentences. After deteding the sentence boundaries each sentence is written in its own line.

→ 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, spokes-Tim Wagner said.
United, a unit of UAL, said the increase took effect Thursday night and applies to most routes where its completes against discount carriers, such as Chicago to Dallas and Atlanta and Denver to San Francisco, Los Angeles and New York.

b) Lemmatization :

Lemmatization usually refers to doing things properly with the use of a vocabulary and morphological analysis of words, normally aiming to remove inflectional endings only and to return the base or dictionary form of a word, which is known as the lemma.

→ CHICAGO(AP) - Citing high fuel price, United Airlines say Friday it have increase fare by $6 per round trip on flight to some city also serve by lower-cost carrier. American Airlines, a unit AMR, immediately match

Scanned by CamScanner

the move, spokesman Tim Wagner say. United, a unit of UAL, say the increase take effect Thursday night and applies to most route where it complete against discount carrier, such a chicago to Dallas and Atlanta and Denver to San Francisco, Los Angeles and New York.

C. POS tagging / Chunking:

Part-of-Speech Tagger (POS Tagger) is a piece of software that reads text in some language and assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc., although generally computational applications use more fine-grained POS tags like 'noun-plural'.

→ 1. [('CHICAGO (AP', 'NNP'), ('(AP)', 'NNP'), ('–', ':'), ('citing', 'VBG'), ('high', 'JJ'), ('fuel', 'NN'), ('prices', 'NNS'), (',', ','), ('United', 'NNP'), ('Airlines', 'NNPS'), ('said', 'VBD'), ('Friday', 'NNP'), ('it', 'PRP'), ('has', 'VBZ'), ('increased', 'VBN'), ('fares', 'NNS'), ('by', 'IN'), ('$', '$'), ('6', 'CD'), ('per', 'IN'), ('round', 'NN'), ('trip', 'NN'), ('on', 'IN'), ('flights', 'NNS'), ('to', 'TO'), ('some', 'DT'), ('cities', 'NNS'), ('also', 'RB'), ('served', 'VBD'), ('by', 'IN'), ('lower-cost', 'JJ'), ('carriers', 'NNS'), ('.', '.'),

2. ('American', 'NNP'), ('Airlines', 'NNPS'), (',', ','), ('a', 'DT'), ('unit', 'NN'), ('AMR', 'NNP'), (',', ','), ('immediately', 'RB'), ('matched', 'VBD'), ('the', 'DT'), ('move', 'NN'), (',', ','), ('spokesman', 'NN'), ('Tim', 'NNP'), ('Wagner', 'NNP'), ('said', 'VBD'), ('.', '.'),

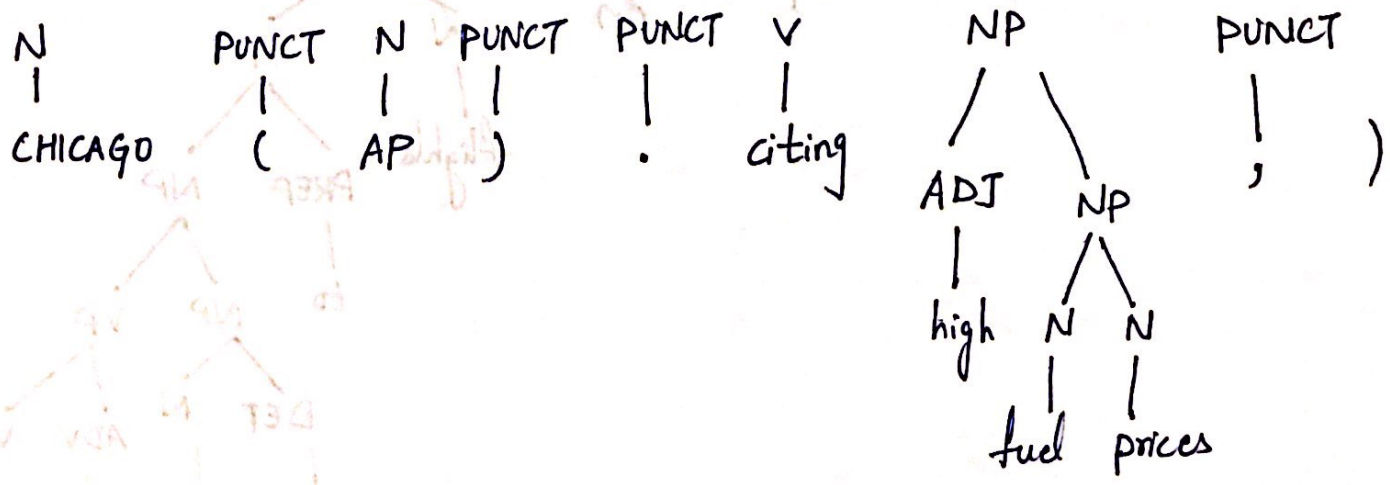
3. ('United', 'NNP'), (',', ','), ('a', 'DT'), ('unit', 'NN'), ('of', 'IN'),

Scanned by CamScanner

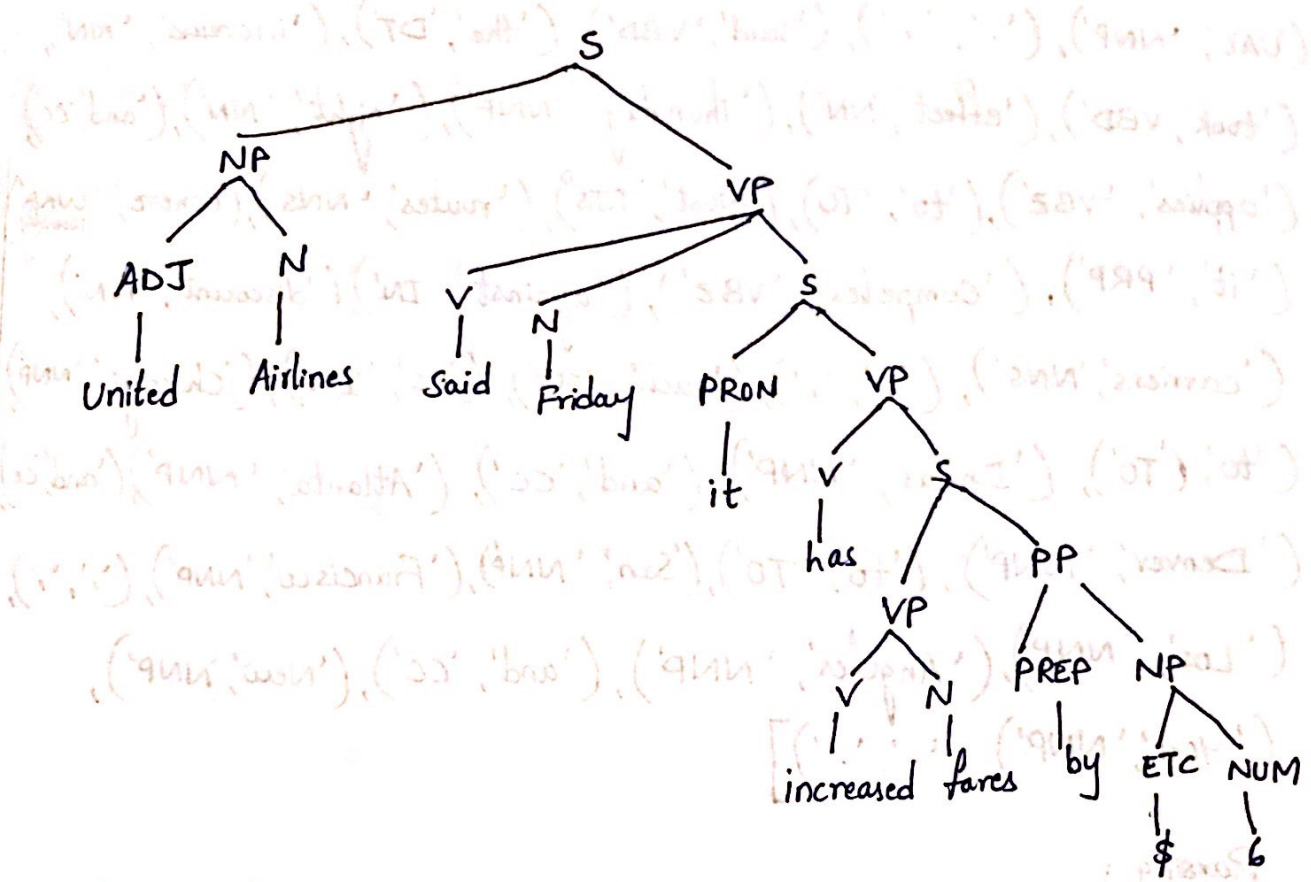
('UAL', 'NNP'), (',', ','), ('said', 'VBD'), ('the', 'DT'), ('increase', 'NN'), ('took', 'VBD'), ('effect', 'NN'), ('Thursday', 'NNP'), ('night', 'NN'), ('and', 'CC'), ('applies', 'VBZ'), ('to', 'TO'), ('most', 'JJS'), ('routes', 'NNS'), ('where', 'WRB'), ('it', 'PRP'), ('competes', 'VBZ'), ('against', 'IN'), ('discount', 'NN'), ('carriers', 'NNS'), (',', ','), ('such', 'JJ'), ('as', 'IN'), ('Chicago', 'NNP'), ('to', 'TO'), ('Dallas', 'NNP'), ('and', 'CC'), ('Atlanta', 'NNP'), ('and', 'CC'), ('Denver', 'NNP'), ('to', 'TO'), ('San', 'NNP'), ('Francisco', 'NNP'), (',', ','), ('Los', 'NNP'), ('Angeles', 'NNP'), ('and', 'CC'), ('New', 'NNP'), ('York', 'NNP'), ('.', '.')]

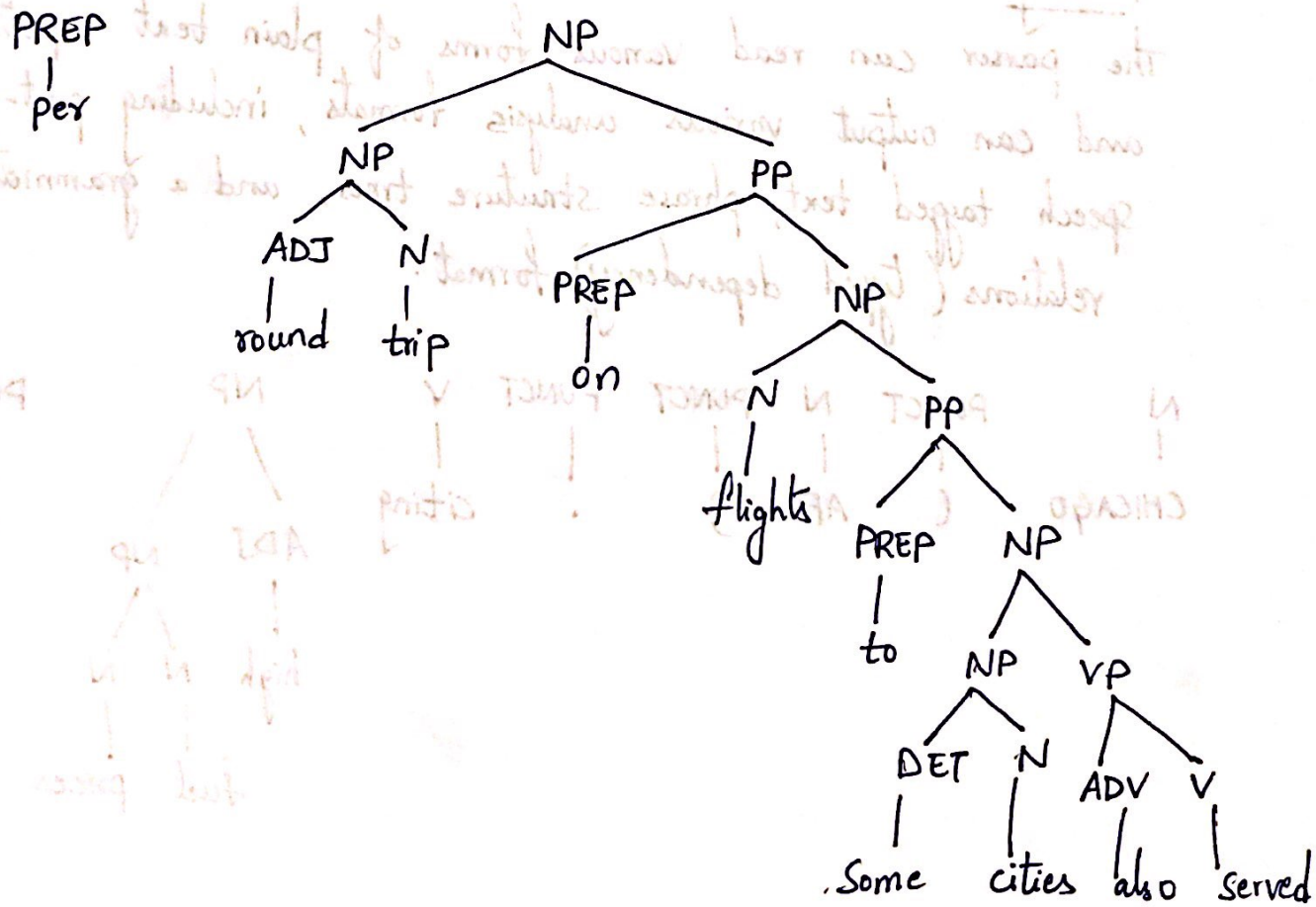
d. Parsing :

The parser can read various forms of plain text input and can output various analysis formats, including part-of-speech tagged text, phrase structure trees, and a grammatical relations (typed dependency) format.

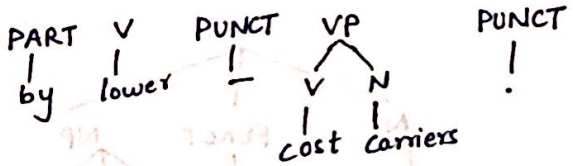


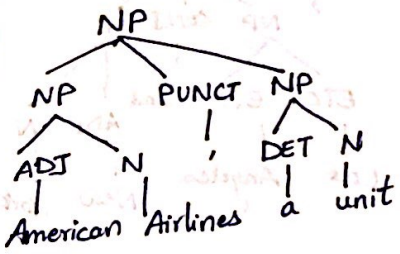
Scanned by CamScanner

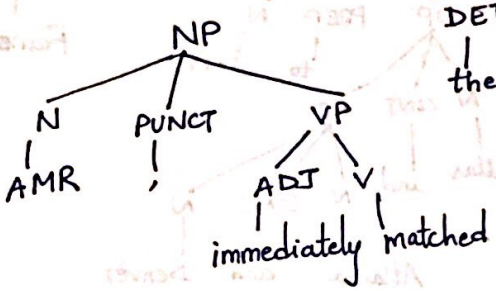


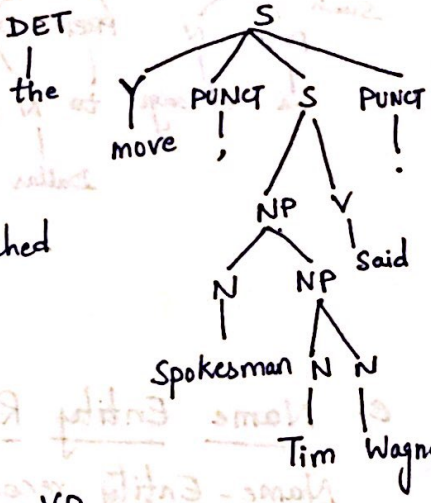


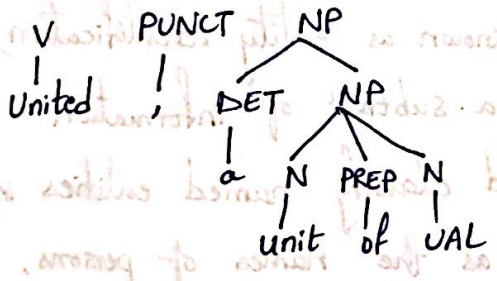
Scanned by CamScanner

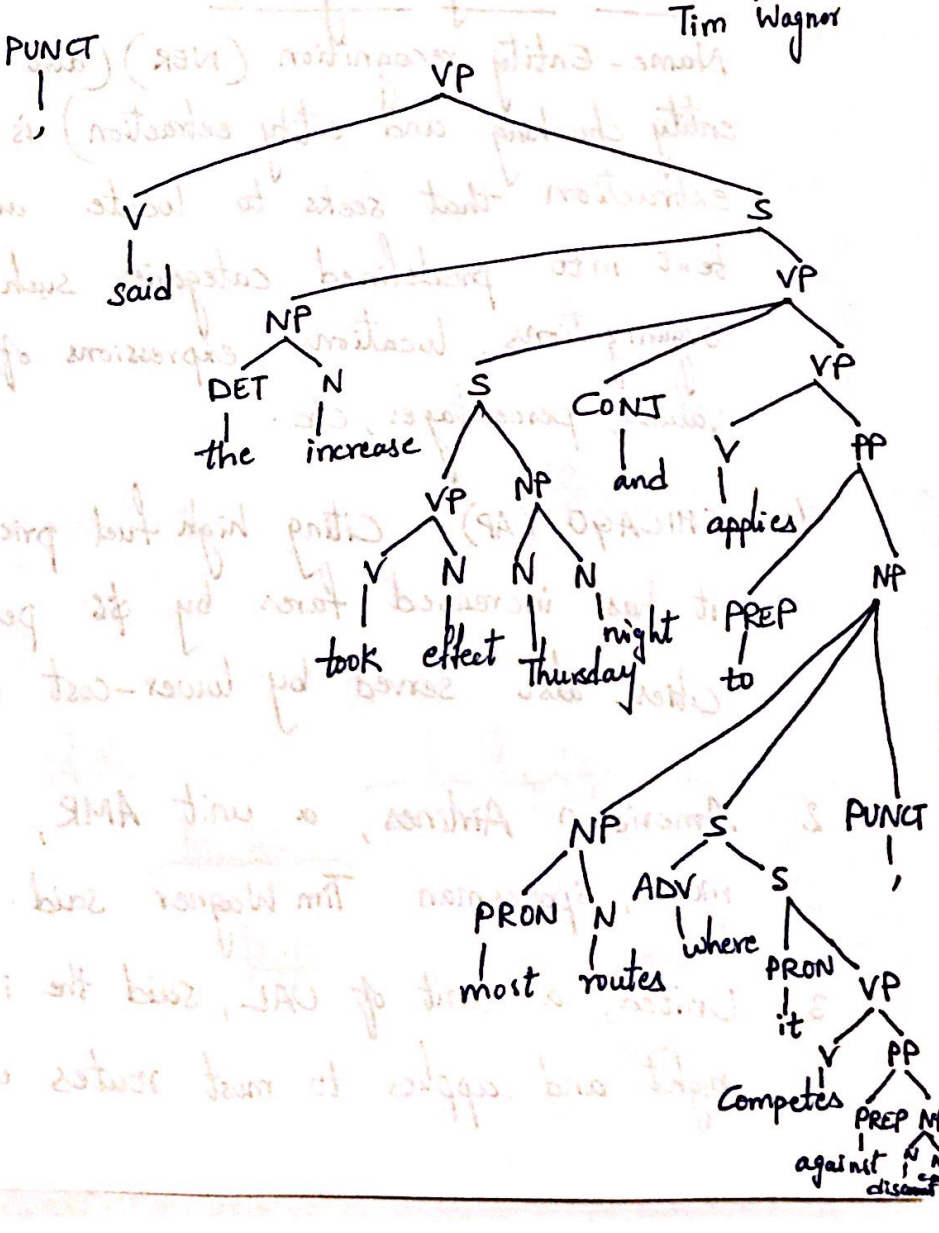




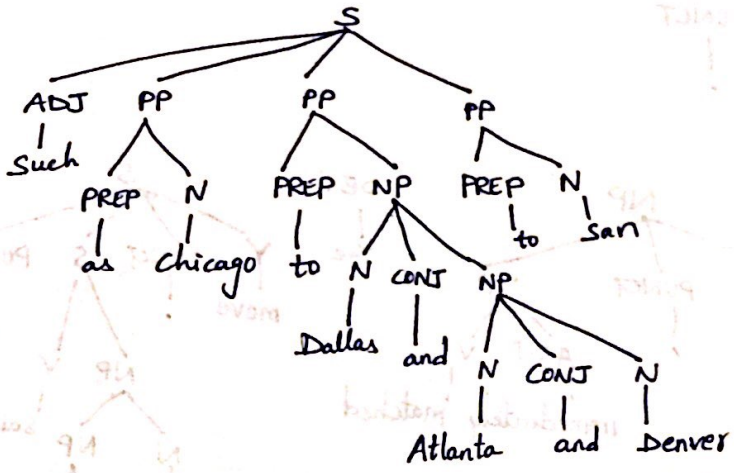


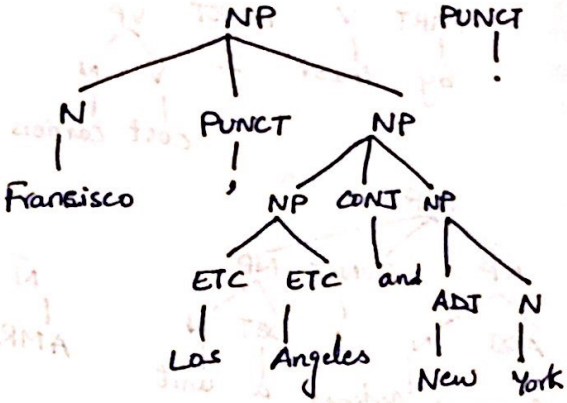






Scanned by CamScanner





e. Name Entity Recognition

Name-Entity recognition (NER) (also known as entity identification, entity chunking and entity extraction) is a subtask of information extraction that seeks to locate and classify named entities in text into predefined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.

1. 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.
   (Location, Org, organization, Date)

2. American Airlines, a unit AMR, immediately matched the move, spokesman Tim Wagner said.
   (Organization, Org, Person)

3. United, a unit of UAL, said the increase took effect Thursday night and applies to most routes where it competes against discount
   (Org, Org, Date, Time)

Scanned by CamScanner

Carriers, such as Chicago (Location) to Dallas (Loc) and Atlanta (Location) and Denver (Loc) to San Francisco (Location), Los Angeles (Location) and New York (Location).

f. Co-reference Resolution:

Co-reference occurs when two or more expressions in a text refer to the same person or thing; they have the same referent, e.g. Bill said he would come; the proper noun Bill and the pronoun he refer to the same person, namely to Bill. Coreference is the main concept underlying binding phenomena in the field of syntax.

→

1. CHICAGO (AP) – Citing high fuel prices, United Airlines (Mention) said Friday it (M) has increased fares by $6 per round trip on flights to some cities also served by lower-cost carriers.

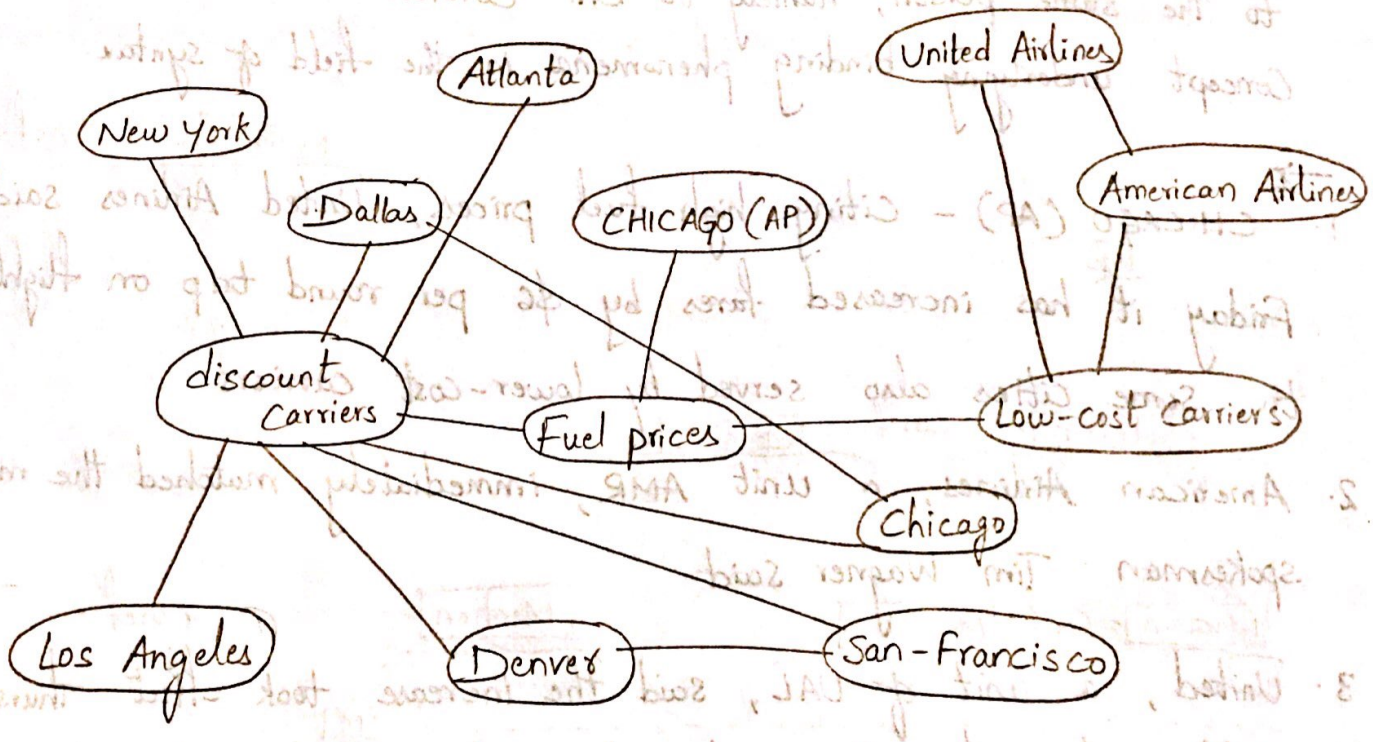
2. American Airlines (Mention), a unit AMR (Mention), immediately matched the move, spokesman Tim Wagner said.

3. United (Mention), a unit of UAL, said the increase (Mention) took effect Thursday night and applies to most routes where it (M) competes against discount carriers, such as Chicago to Dallas and Atlanta and Denver to San Francisco, Los Angeles and New York.

Scanned by CamScanner

2. Knowledge Graph:

The Knowledge Graph is a knowledge base used by the Google to enhance its search engine's search results with semantic-search information gathered from a wide variety of sources. the Google Knowledge Graph search API is a tool that helps users find entities with in the Google Knowledge Graph.



Scanned by CamScanner