CS5560 Knowledge Discovery and Management

In-Class-Exercise (ICE-1A)

June 5, 2017

Student Names: Sujitha Puthana

Class ID : 24

We are supposed to build a knowledge graph for the following input (text data).

1. Describe your knowledge about knowledge graph.
2. Why do we want to build such a knowledge graph?
3. What steps are required? Show your own workflow for this task.
4. What are the challenges?
5. Draw a knowledge graph for the given data.

> 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.

① Knowledge Graph

-> Knowledge Graph is the graphical representation of the text after sorting through the Natural language processing Algorithm.

-> Knowledge Graph is highly used by google for its search engine. using the semantic search.

② Why knowledge Graph?

-> Knowledge Graph helps in simplifying the search engine.

-> Huge amount of data (text) is constructed to a Knowledge Graph. Linking of these multiple knowledge graph, helps in decision making.

-> Reduces the redundancy, making it cost effective.

(3) Steps for Knowledge Graph :-

Step 1 - Natural Language Processing

Annotation → Lemmatization → POS Tagging → Named Entity Recognization

CHICAGO	→ CHICAGO	→ NNP	→ LOCATION
LRB	→ -lrb-	→ - LRB -	
AP	→ AP	→ NER	→ ORGANIZATION
-RRB-	→ -rrb-	→ -RRB-	
liting	→ lite	→ VBG	
high	→ high	→ JJ	
fuel	→ fuel	→ NN	
prices	→ prices	→ NNS	
,	→ ,	→ ,	
United	→ United	→ NNP	→ Organization
Airlines	→ Airlines	→ NNPS	→ Organization
said	→ say	→ VBD	
Friday	→ Friday	→ NNP	→ Date
it	→ it	→ PRP	
has	→ have	→ VBZ	
increased	→ increase	→ VBN	
fares	→ fare	→ NNS	
by	- by	→ IN	
$	- $	→ $	→ Money

Annotation	→ Lemmatization	→ POS tagging	→ Named Entity Recognition
Airlines	→ Airlines	→ NNPS	→ Organization
6	→ 6	→ CD	→ Money
per	→ per	→ IN	
round	→ round	→ NN	
trip	→ trip	→ NN	
On	→ On	→ IN	
flights	→ flight	→ NNS	
to	→ to	→ TO	
some	→ some	→ DT	
cities	→ city	→ NNS	
also	→ also	→ RB	
served	→ serve	→ VBD	
by	→ by	→ IN	
lower-cost	→ lower-cost	→ JJ	
carriers	→ carrier	→ NNS	
.	→ .	→ .	
American	→ American	→ NNP	→ organization
,	→ ,	→ ,	
a	→ a	→ DT	
unit	→ unit	→ NN	
AMR	→ AMR	→ NNP	
,	→ ,	→ ,	

immediately → immediately → RB

matched → match → VBD

the → the → DT

move → move → N V

, → , → ,

Spokesman → spokesman → NN

Tim → Tim → NNP → PERSON

Wagner → Wagner → NNP → PERSON

said → say → VBD

United → United → NNP → Organization

a → a → DT

unit → unit → NN

of → of → IN

UAL → UAL → NNP → Organization

, → , → ,

said → say → VBD

the → the → DT

increase → increase → NN

took → take → VBD

effect → effect → NN

Thursday → Thursday → NNP → DATE

night → night → NN → TIME

and → and → CC

applies → apply → VBZ

to → to → TO

Annotation → Lemmatization → POS tagging → Named Entity Recognization

most → most → JJS

routes → route → NNS

where → where → WRB

it → it → PRP

competes → compete → VBZ

against → against → IN

discount → discount → NN

carriers → carrier → NNS

, → , → ,

such → such → JJ

as → as → IN

Chicago → Chicago → NNP → Location

to → to → TO

Dallas → Dallas → NNP → Location

and → and → CC

Allbata → Allbata → NNP → Location

And → and → CC

Denver → Denver → NNP → Location

to → to → TO

SanFrancisco → SanFrancisco → NNP → Location

Los Angles → LosAngles → NNP → Location

and → and → CC

New York → New York → NNP → Location

Natural Language Processing refers to the use and ability of systems to process sentences in a natural language such as English, rather than in a specialized artificial computer language such as C++.

Steps involved are:-

1. Tokenization
2. Stemming / Lemmatization
3. Parts of Speech tagging
4. Query Expansion
5. Parsing
6. Topic segmentation Recognisation / Name Entity recognization.
7. Morphological Segmentation (word / sentence)

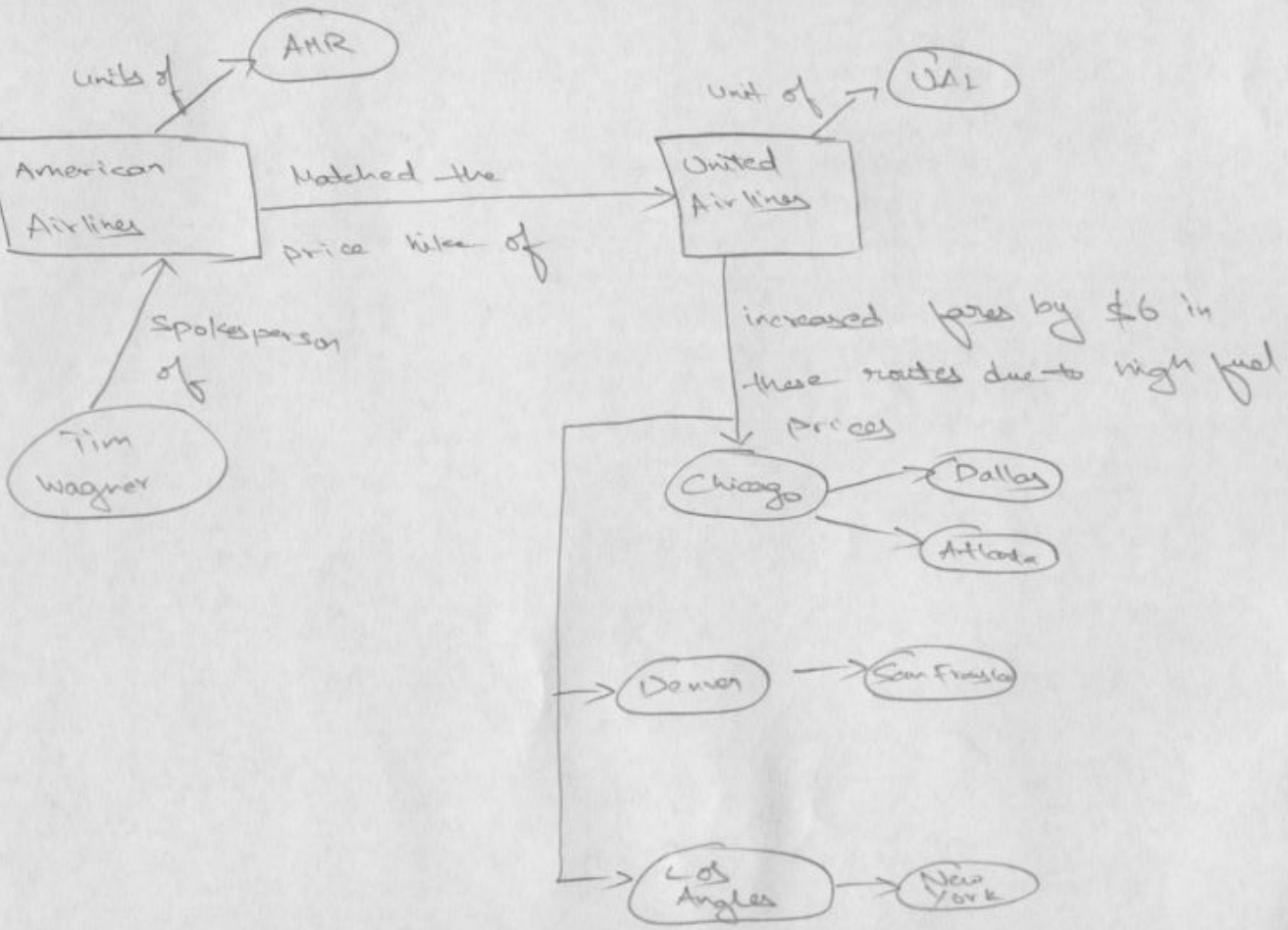
Step 2 - Information retrieval and information extraction

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 Angles and New York.

Step 3 :- Topic Discovery

Organization	People	Places
United Airlines	Tim Wagner	Chicago
American Airlines		Dallas
AMR		Atlanta
UAL		Denver
		San Francisco
		Los Angeles
		New York

Step 4 :- Knowledge Graph Construction :-



(4) Challenges :-

-> Different Knowledge Graph can be created, as there are no specific format to design these Knowledge graph.

-> Knowledge Graph cannot be complete.

-> Accuracy - Cannot judge if the knowledge is accurate.