CS5560 Knowledge Discovery and Management

Problem Set (PS-1B)

June 5, 2017

Name: **Revanth Chakilam**

Class ID: **02**

St ID: **16241583**

a. Sentence detection/Tokenization:

we will assign the numbers for all the words in the given paragraph by dividing in each sentences.

CHICAGO-1, -LRB--2, AP-3, -RRB--4, ---5, Citing-6, high-7, fuel-8, prices-9, ,-10, United-11, Airlines-12, said-13,Friday-14,it-15,has-16,increased-17,fares-18,by-19,$-20,6-21,per-22,round-23,trip-24,on-25,flights-26,to-27,some-28,cities-29,also-30,served-31,by-32,lower-cost-33,carriers-34,.-35

American-1, Airlines-2, ,-3, a-4,unit-5,AMR-6,,-7,immediately-8,matched-9,the-10,move-11,,-12,spokesman-13,Tim-14,Wagner-15,said-16,.-17

United-1,,-2,a-3,unit-4,of-5,UAL-6,,-7,said-8,the-9,increase-10,took-11,effect-12,Thursday-13,night-14,and-15,applies-16,to-17,most-18,routes-19,where-20,it-21,competes-22,against-23,discount-24,carriers-25,,-26,such-27,as-28,Chicago-29,to-30,Dallas-31,and-32,Atlanta-33,and-34,Denver-35,to-36,San-37,Francisco-38,,-39,Los-40,Angeles-41,and-42,New-43,York-44,.-45

b. Lemmatization

The goal of stemming and lemmatization is to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form

CHICAGO-1:CHICAGO -LRB--2:-lrb- AP-3:AP -RRB--4:-rrb- ---5:-- Citing-6:cite high-7:high fuel-8:fuel prices-9:price ,-10:, United-11:United Airlines-12:Airlines said-13:say Friday-14:Friday it-15:it has-16:have increased-17:increase fares-18:fare by-19:by $-20:$ 6-21:6 per-22:per round-23:round trip-24:trip on-25:on flights-26:flight to-27:to some-28:some cities-29:city also-30:also served-31:serve by-32:by lower-cost-33:lower-cost carriers-34:carrier .-35:.

American-1: American Airlines-2: Airlines ,-3:, a-4:a unit-5:unit AMR-6:AMR ,-7:, immediately-8:immediately matched-9:match the-10:the move-11:move ,-12:, spokesman-13:spokesman Tim-14:Tim Wagner-15:Wagner said-16:say .-17:.

United-1:United ,-2:, a-3:a unit-4:unit of-5:of UAL-6:UAL ,-7:, said-8:say the-9:the increase-10:increase took-11:take effect-12:effect Thursday-13:Thursday night-14:night and-15:and applies-16:apply to-17:to most-18:most routes-19:route where-20:where it-21:it competes-22:compete against-23:against discount-24:discount carriers-25:carrier ,-26:, such-27:such as-28:as Chicago-29:Chicago to-30:to Dallas-31:Dallas and-32:and Atlanta-33:Atlanta and-34:and Denver-35:Denver to-36:to San-37:San Francisco-38:Francisco ,-39:, Los-40:Los Angeles-41:Angeles and-42:and New-43:New York-44:York .-45:.

A Part-Of-Speech Tagger - POS Tagger reads text in some language assigns parts of speech to each word such as noun, verb, adjective, etc., although generally computational applications use more fine-grained POS tags like 'noun-plural'.

CHICAGO-1:NNP -LRB--2:-LRB- AP-3:NNP -RRB--4:-RRB- ---5:: Citing-6:VBG high-7:JJ fuel-8:NN prices-9:NNS ,-10:, United-11:NNP Airlines-12:NNPS said-13:VBD Friday-14:NNP it-15:PRP has-16:VBZ increased-17:VBN fares-18:NNS by-19:IN $-20:$ 6-21:CD per-22:IN round-23:NN trip-24:NN on-25:IN flights-26:NNS to-27:TO some-28:DT cities-29:NNS also-30:RB served-31:VBD by-32:IN lower-cost-33:JJ carriers-34:NNS .-35:.

American-1: NNP Airlines-2: NNPS ,-3:, a-4:DT unit-5:NN AMR-6:NNP ,-7:, immediately-8:RB matched-9:VBD the-10:DT move-11:NN ,-12:, spokesman-13:NN Tim-14:NNP Wagner-15:NNP said-16:VBD .-17:.

United-1:NNP ,-2:, a-3:DT unit-4:NN of-5:IN UAL-6:NNP ,-7:, said-8:VBD the-9:DT increase-10:NN took-11:VBD effect-12:NN Thursday-13:NNP night-14:NN and-15:CC applies-16:VBZ to-17:TO most-18:JJS routes-19:NNS where-20:WRB it-21:PRP competes-22:VBZ against-23:IN discount-24:NN carriers-25:NNS ,-26:, such-27:JJ as-28:IN Chicago-29:NNP to-30:TO Dallas-31:NNP and-32:CC Atlanta-33:NNP and-34:CC Denver-35:NNP to-36:TO San-37:NNP Francisco-38:NNP ,-39:, Los-40:NNP Angeles-41:NNP and-42:CC New-43:NNP York-44:NNP .-45:.

A natural language parser is a program that works out the grammatical structure of sentences, for instance, which groups of words go together (as "phrases") and which words are the subject or object of a verb

 (ROOT (NP (NP (NP (NNP CHICAGO)) (PRN (-LRB- -LRB-) (NP (NNP AP)) (-RRB- -RRB-))) (: --) (S (S (VP (VBG Citing) (NP (JJ high) (NN fuel) (NNS prices)))) (, ,) (NP (NNP United) (NNPS Airlines)) (VP (VBD said) (NP-TMP (NNP Friday)) (SBAR (S (NP (PRP it)) (VP (VBZ has) (VP (VBN increased) (NP (NNS fares)) (PP (IN by) (NP (NP ($ $) (CD 6)) (PP (IN per) (NP (NN round) (NN trip)))))) (PP (IN on) (NP (NNS flights))) (PP (TO to) (NP (NP (DT some) (NNS cities)) (VP (ADVP (RB also)) (VBD served) (PP (IN by) (NP (JJ lower-cost) (NNS carriers))))))))))))) (. .)))

(ROOT (S (S (NP (NP (NNP American) (NNPS Airlines)) (, ,) (NP (NP (DT a) (NN unit)) (NP (NNP AMR))) (, ,)) (ADVP (RB immediately)) (VP (VBD matched) (NP (DT the) (NN move)))) (, ,) (NP (NN spokesman) (NNP Tim) (NNP Wagner)) (VP (VBD said)) (. .)))

 (ROOT (S (NP (NP (NNP United)) (, ,) (NP (NP (DT a) (NN unit)) (PP (IN of) (NP (NNP UAL)))) (, ,)) (VP (VBD said) (SBAR (S (NP (DT the) (NN increase)) (VP (VP (VBD took) (NP (NN effect)) (NP-TMP (NNP Thursday) (NN night))) (CC and) (VP (VBZ applies) (PP (TO to) (NP (JJS most) (NNS routes))) (SBAR (WHADVP (WRB where)) (S (NP (PRP it)) (VP (VBZ competes) (PP (IN against) (NP (NP (NN discount) (NNS carriers)) (, ,) (PP (JJ such) (IN as) (NP (NP (NP (NNP Chicago)) (PP (TO to) (NP (NNP Dallas) (CC and) (NNP Atlanta)))) (CC and) (NP (NP (NNP Denver)) (PP (TO to) (NP (NP

(NNP San) (NNP Francisco)) (, ,) (NP (NNP Los) (NNP Angeles)) (CC and) (NP (NNP New) (NNP York)))))))))))))))))) (. .)))

CHICAGO-1:LOCATION -LRB--2:O AP-3:ORGANIZATION -RRB--4:O ---5:O Citing-6:O high-7:O fuel-8:O prices-9:O ,-10:O United-11:ORGANIZATION Airlines-12:ORGANIZATION said-13:O Friday-14:DATE it-15:O has-16:O increased-17:O fares-18:O by-19:O $-20:MONEY 6-21:MONEY per-22:O round-23:O trip-24:O on-25:O flights-26:O to-27:O some-28:O cities-29:O also-30:O served-31:O by-32:O lower-cost-33:O carriers-34:O .-35:O

American-1: ORGANIZATION Airlines-2: ORGANIZATION,-3:O a-4:O unit-5:O AMR-6:ORGANIZATION ,-7:O immediately-8:O matched-9:O the-10:O move-11:O ,-12:O spokesman-13:O Tim-14:PERSON Wagner-15:PERSON said-16:O .-17:O

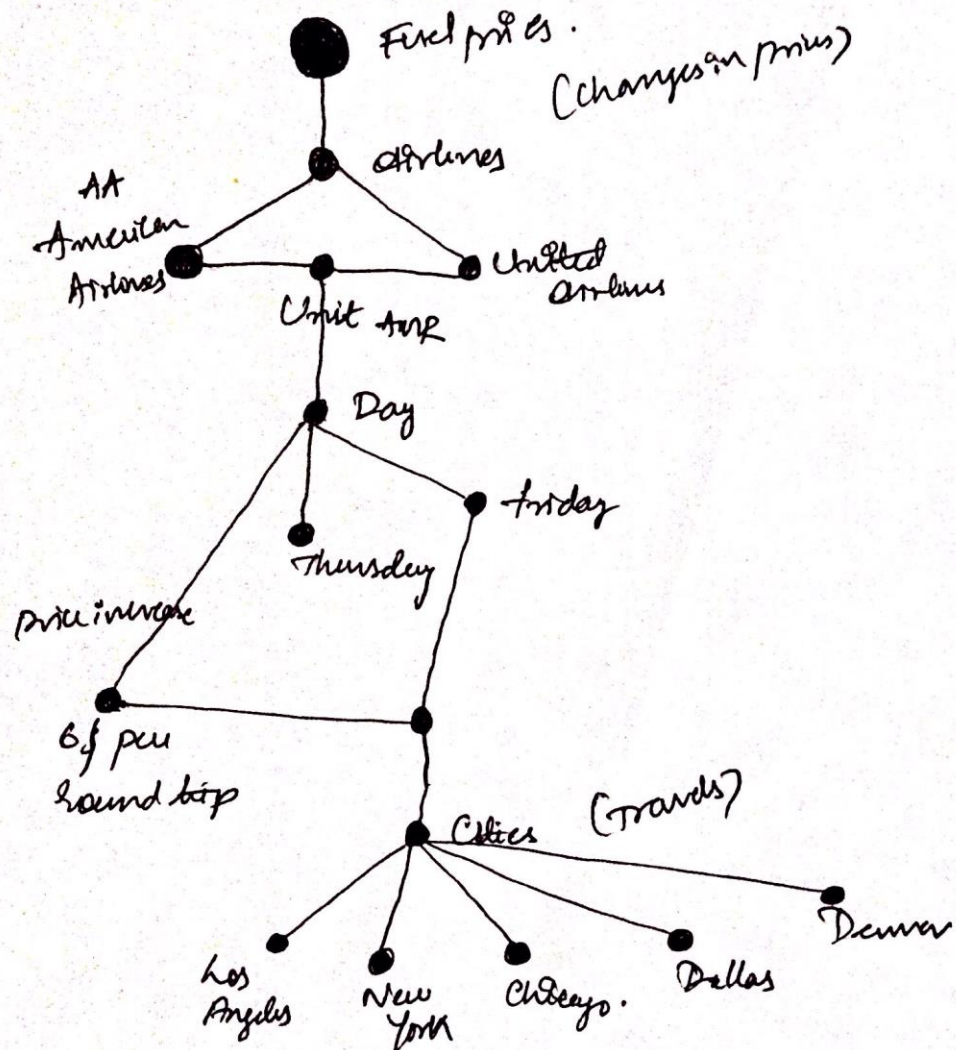United-1:ORGANIZATION ,-2:O a-3:O unit-4:O of-5:O UAL-6:ORGANIZATION ,-7:O said-8:O the-9:O increase-10:O took-11:O effect-12:O Thursday-13:DATE night-14:TIME and-15:O applies-16:O to-17:O most-18:O routes-19:O where-20:O it-21:O competes-22:O against-23:O discount-24:O carriers-25:O ,-26:O such-27:O as-28:O Chicago-29:LOCATION to-30:O Dallas-31:LOCATION and-32:O Atlanta-33:LOCATION and-34:O Denver-35:LOCATION to-36:O San-37:LOCATION Francisco-38:LOCATION ,-39:O Los-40:LOCATION Angeles-41:LOCATION and-42:O New-43:LOCATION York-44:LOCATION .-45:O

[CHAIN1-["AP" in sentence 1], CHAIN2-["United Airlines" in sentence 1, "it" in sentence 1], CHAIN3-["CHICAGO -LRB- AP -RRB- -- 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 ." in sentence 1], CHAIN4-["high fuel prices" in sentence 1], CHAIN5-["Friday" in sentence 1], CHAIN7-["fares" in sentence 1], CHAIN8-["round trip" in sentence 1], CHAIN9-["flights" in sentence 1], CHAIN10-["some cities also served by lower-cost carriers" in sentence 1], CHAIN11-["lower-cost carriers" in sentence 1], CHAIN12-["American Airlines" in sentence 2, "a unit AMR" in sentence 2], CHAIN13-["AMR" in sentence 2], CHAIN14-["American Airlines , a unit AMR" in sentence 2], CHAIN16-["the move" in sentence 2], CHAIN17-["spokesman Tim Wagner" in sentence 2], CHAIN18-["United" in sentence 3, "a unit of UAL" in sentence 3], CHAIN19-["UAL" in sentence 3], CHAIN20-["Dallas" in sentence 3], CHAIN21-["Atlanta" in sentence 3], CHAIN22-["San Francisco" in sentence 3], CHAIN23-["Los Angeles" in sentence 3], CHAIN24-["New York" in sentence 3], CHAIN25-["United , a unit of UAL" in sentence 3], CHAIN27-["the increase" in sentence 3, "it" in sentence 3], CHAIN28-["Thursday night" in sentence 3], CHAIN29-["most routes" in sentence 3], CHAIN31-["discount carriers , such as Chicago to Dallas and Atlanta and Denver to San Francisco , Los Angeles and New York" in sentence 3], CHAIN32-["discount carriers" in sentence 3], CHAIN33-["Chicago to Dallas and Atlanta and Denver to San Francisco , Los Angeles and New York" in sentence 3], CHAIN34-["Chicago to Dallas and Atlanta" in sentence 3], CHAIN35-["Dallas and Atlanta" in sentence 3], CHAIN36-["Denver to San Francisco , Los Angeles and New York" in sentence 3], CHAIN37-["San Francisco , Los Angeles and New York" in sentence 3]]

5)

(fig) knowledge graph for the example given