The University of Texas at Dallas
CS 6320
Natural Language Processing
Spring 2021
Instructor: Dr. Sanda Harabagiu
Grader/ Teaching Assistant: Maxwell Wienzierl

Homework 3: 100 points
Issued April 14, 2021
Due May 3, 2021 before midnight

PROBLEM 1: Semantic Role Labeling (SRL) (**40 points**)
Considering the following four sentences:

S1: Sales of the company to return to normalcy.
S2: The new products and services contributed to increase revenue.
S3: Dow falls as recession indicator flashed red and economical worries continue through the month.
S4: Figure skater lands historic quadruple jump in senior international competition at the 2019 World Figure Skating Championships on Day 3 but could only clinch a silver medal.

Identify by hand the predicates and their arguments against ProbBank definitions, available at:
http://verbs.colorado.edu/verb-index/index.php
Indicate clearly the span of the arguments on each sentence, by marking them as:

Sales of [the company]$_{Arg???}$ to return to normalcy.

PROBLEM 2: Information Extraction (**60 points**)
a/ Identify manually the Named Entities in the following short text: (**10 points**)

Solange Knowles will not be performing at this year's Coachella Music and Arts Festival after all. The music festival's official Twitter account announced the news Sunday night. Knowles released her new album, "When I Get Home," last month. An accomplished singer, songwriter, dancer, DJ and style guru, Knowles graced the Coachella stage last year when she joined her big sister, Beyoncé, during her historic, headlining tour.
Ariana Grande, Childish Gambino, and Tame Impala are headlining the festival this year. Coachella takes place annually over two weekends in April in Indio, California.

Use only the following Name Entity Categories and their respective acronyms:
PERSON (PER), LOCATION (LOC), ENTERTAINMENT-EVENT (ENT), ORGANIZATION (ORG), ALBUM-TITLE (ALB).

The document was posted on April 14th, 2019.

Use the following format to mark the named entities that you identify:

Citing high fuel prices, [ORG **United Airlines**] said [TIME **Friday**] it has increased fares by [MONEY **$6**] per round trip on flights to some cities also served by lower-cost carriers. [ORG **American Airlines**], a unit of [ORG **AMR Corp.**], immediately matched the move, spokesman [PER **Tim Wagner**] said. [ORG **United**], a unit of [ORG **UAL Corp.**], said the increase took effect [TIME **Thursday**] and applies to most routes where it competes against discount carriers, such as [LOC **Chicago**] to [LOC **Dallas**] and [LOC **Denver**] to [LOC **San Francisco**].

State how many mentions of Named entities you have found and note how many entities were Persons, how many were Locations, how many were Entertainment Events, how many were Organizations and how many were Album Titles.

b/ (5 points) Use the IOB notation to indicate the labels that you would assign to the text from Problem 2.a. To indicate the labels, use the following table formatting for each sentence:

Words	IOB Label
American	B-ORG
Airlines	I-ORG
,	O
a	O
unit	O
of	O
AMR	B-ORG
Corp.	I-ORG
,	O
immediately	O
matched	O
the	O
move	O
,	O
spokesman	O
Tim	B-PER
Wagner	I-PER
said	O
.	O

Generate a different table for each sentence in the text, e.g. Table 1 for sentence 1, Table 2 for sentence 2 etc.

c/ Annotate manually the temporal expressions in the text from Problem 2.a, using the following format (**5 points**):

> A fare increase initiated <TIMEX3>last week</TIMEX3> by UAL Corp's United Airlines was matched by competitors over <TIMEX3>the weekend</TIMEX3>, marking the second successful fare increase in <TIMEX3>two weeks</TIMEX3>.

d/ (**10 points**) Perform temporal normalization of the temporal expressions you have identified in the text from Problem 2.a and produce an annotation of the normalized temporal expressions in the text.

e/ (**25 points**) Identify in the text from Problem 2.a all the relations of the following types:

TYPE 1: PERSON → *Entertainer* → ENTERTAINMENT-EVENT (PER → *Entertainer* → ENT)

TYPE 2: PERSON → *Family* → PERSON (PER → *Family* → PER)

TYPE 3: PERSON → *Involved-in* → ALBUM-TITLE (PER → *Involved-in* → ALB)

TYPE 4: ENTERTAINMENT-EVENT → *Happens-in* → LOCATION (ENT → *Happens-in* → LOC)

TYPE 5: ENTERTAINMENT-EVENT → *Communicates-through* → ORGANIZATION (ENT → *Communicates-through* → ORG)

When presenting the relations that you manually discover, use the following format:

[Tim Warner]PER → Spokesman → [American Airlines]ORG

Indicate how many relations of each type that you have discovered.

f/ (**5 points**) Provide examples of five classes of feature you would use to discover the five types of relations you have identified manually. Select one of your relations and show the feature values that you extract from the text from Problem 2.a for 2 of the five feature you selected. Discuss the classifier you would use for recognizing automatically these relations.