CS-5340/6340, Written Assignment #3
DUE: Friday, November 18, 2016 by 11:00pm

1. (20 pts) Identify the thematic role of each noun phrase in the sentences below. Choose from the following set of thematic roles: *agent, co-agent, theme, co-theme, instrument, recipient, experiencer, beneficiary, origin, destination, path*

   (a) John cleared the sidewalk with his snow shovel.

   (b) A raffle was held for the local charity.

   (c) Joe teamed up with Susan to learn how to ski.

   (d) Tom was given a bike.

   (e) The dress was purchased by Lisa with matching shoes.

   (f) Julie mailed her grandmother a letter.

   (g) George hopes that his car can be fixed.

   (h) The image was created with Adobe Photoshop by a cartoonist.

2. (14 pts) Consider the following context vectors:

$word1$: <9 2 8 0>
$word2$: <6 7 4 5>
$word3$: <1 3 9 2>

Compute the similarity between pairs of the vectors above using the similarity metrics stated below. Please show all your work!

(a) Similarity($word1$, $word2$) using Manhattan Distance.

(b) Similarity($word2$, $word3$) using Manhattan Distance.

(c) Similarity($word1$, $word2$) using Jaccard Similarity.

(d) Similarity($word2$, $word3$) using Jaccard Similarity.

(e) Similarity($word1$, $word2$) using Cosine Similarity.

(f) Similarity($word2$, $word3$) using Cosine Similarity.

3. (30 pts) This question relates to the Collins & Singer bootstrapping method for named entity recognition. The predicate Contains(w) is satisfied if an NP or Context includes the word w. The table below contains NP/Context pairs extracted from an imaginary text corpus. Please treat all words as being case-insensitive, for example "city", "City", and "CITY" should all be considered to be the same word.

NP	**CONTEXT**	**CLASS**
Apple tablet	mobile computer	PRODUCT
British tablet	computer tech	PRODUCT
Microsoft tablet	mobile computer	PRODUCT
Apple	tech giant	COMPANY
Apple Corporation	tech leader	COMPANY
Microsoft	leader	COMPANY
Microsoft Corporation	tech giant	COMPANY
British Tablet Corporation	mobile computer giant	COMPANY

(a) Using the Contains(w) predicate, list all of the rules that would be generated from the NPs in the table above, and compute the probability for each rule. Each rule should be of the form:

If Contains(<word>) → <class>

Leave the probabilities in fractional form! Do not do any iterative bootstrapping, assume this is the very first iteration of the algorithm.

(b) Using the Contains(w) predicate, list all of the rules that would be generated from the Contexts in the table above, and compute the probability of each rule. Each rule should be of the form:

If Contains(<word>) → <class>

Leave the probabilities in fractional form! Do not do any iterative bootstrapping, assume this is the very first iteration of the algorithm.

4. (18 pts) Assume that an imaginary thesaurus contains the following word lists for the categories FISH and MUSIC.

   FISH: eel, ray, trout
   MUSIC: guitar, symphony, tenor

   The box below shows 7 sentences, which you should use as the text corpus for this question. This text corpus contains exactly 60 words. Please treat all words as being case-insensitive, for example "city", "City", and "CITY" should all be considered to be the same word.

   S1: Trout can be found in many Utah rivers.
   S2: The Utah Symphony hired a new bass player.
   S3: John Smith plays an electric bass guitar.
   S4: A soprano and bass will sing with the symphony.
   S5: Ray Charles is learning to play the electric bass guitar.
   S6: The Utah aquarium displays an electric eel, bass, and guitar fish.
   S7: The Utah fisherman caught a river trout.

   Compute the following salience values as defined by Yarowsky's word sense disambiguation algorithm. You should assume that the context window for a word spans the entire sentence containing the word but does not cross sentence boundaries. Please show all your work!

   (a) salience(Utah, FISH)

   (b) salience(electric, FISH)

   (c) salience(bass, FISH)

   (d) salience(Utah, MUSIC)

   (e) salience(electric, MUSIC)

   (f) salience(bass, MUSIC)

5. (18 pts) Consider the following short story:

> *John Smith went to the store to buy 10 oranges. At the market John saw Mary, his neighbor. She waved at him and smiled. He tried to wave back but was holding too many groceries and he nearly dropped them. John laughed at himself and Mary laughed too. It so happens that Mary also loves oranges, so John gave Mary some of his oranges. Mary was amazed by their size and thanked him. Mary then found her husband, George, who was in the next aisle and they went home. Coincidentally, it seems that George had purchased oranges too!*

For all questions below, if multiple instances of the same word or phrase fit the criteria, then list each instance separately (e.g., answer "John, John" if two different instances of "John" in the story fit the criteria). If there are no instances that match the criteria, then answer NONE.

(a) List all possible noun phrase (NP) antecedents for the first instance of "He" (underlined) that satisfy gender agreement.

(b) List all possible noun phrase (NP) antecedents for "them" (underlined) that satisfy animacy agreement.

(c) List all possible noun phrase (NP) antecedents for "them" (underlined) that satisfy number agreement.

(d) List all of the reflexive pronouns in the story.

(e) List all of the possessive pronouns in the story.

(f) List all of the pleonastic pronouns in the story.

(g) List all instances of appositives in the story.

ELECTRONIC SUBMISSION INSTRUCTIONS
(a.k.a. "What to turn in and how to do it")

Your written assignment must be in .pdf format. Please do not turn in .doc or .docx files ... convert them to .pdf format before submitting them!

To submit this assignment, the CADE provides a web-based facility for electronic handin, which can be found here:

https://webhandin.eng.utah.edu/

Or you can log in to any of the CADE machines and issue the command:

handin cs5340 written3 <filename>

Please name your file: YourName-written3.pdf (e.g., EllenRiloff-written3.pdf)

HELPFUL HINT: you can get a listing of the files that you've already turned in via electronic submission by using the 'handin' command without giving it a filename. For example:

handin cs5340 written3

will list all of the files that you've turned in thus far. If you submit a new file with the same name as a previous file, the new file will overwrite the old one.