

CS-5340/6340, Written Assignment #3
DUE: TBD, 2018 by 11:59pm

1. (20 pts) Given below is a paragraph of text. Every mention in the text is assigned a label. The mention is marked using the braces [and], and the label is assigned with a / tag.

Perform coreference resolution on this piece of text. Using the **mention-pair model**, list all the pairs that will be generated. For the positive instances, assign the pair a value of 1. For the negative instances, assign a value of 0.

[It/M1] was supposed to be [[Jennifer Lopez/M2]s day/M3] off. Cue [visions/M4] of [her/M5] lounging by her [infinity pool/M6] in [Bel-Air/M7], [friends/M8] hanging, [tunes/M9] turned up. Instead, [Lopez/M10], the [multihyphenate performer/M11], [producer/M12] and [branding maven/M13], held [a half-dozen business meetings/M14] in [[her/M15] home [here/M16]/M17], from [early morning/M18] until [sundown/M19], on [ambitious ventures/M20] ranging from [real estate/M21] to [fitness/M22].

Pair	Value
add your entries here
add your entries here
add your entries here

2. (20 pts) Assume you are working on a relation extraction task to extract professors and their field of study. Your corpus of text is campus media articles. Assume that you are applying a bootstrapping approach with the given seed tuples. Your system extracts the left context, right context, and the middle context (with a max window of 4 words).

Seed Tuples: (Steve Marschner, computer science), (Brian Chabot, ecology and evolutionary biology), (Daniel Schwarz, English Literature)

Some sentences from the text:

Prof. William Jacobson, law, stated in 2015 that he found ...

Prof. Emeritus Brian Chabot, ecology and evolutionary biology, donated to Tracy Mitrano's ...

The New School psychology professor and New University in Exile Consortium founder Arien Mack said ...

Prof. Daniel Schwarz, English Literature who this year celebrated 50 years ...

Daniel Schwarz (born May 12, 1941) is Frederick J. Whiton Professor of **English Literature** and Stephen H. Weiss Presidential Fellow ...

Daniel Schwarz is Frederic J. Whiton Professor of **English Literature** and Stephen H. Weiss Presidential Fellow ...

Daniel Schwarz '68 AM, PhD '68, the Frederic J. Whiton Professor of **English Literature** at Cornell, whos been teaching there ...

Prof. Chekitan Dev, hotel administration, doesn't like to just teach and research in Ithaca.

Prof. Steve Marschner '98, computer science, begged his parents for an Apple II in fourth grade so he could write programs on the 8-bit computer.

... said Kyla Schuller, an associate professor in the Department of Women's and Gender Studies.

Professor Sang-Hyuk Lee of the Department of Physics and Astronomy, and his team, recently received the \$1.5 million grant from the Department of Energy.

Steve Marschner joined the faculty at Cornell in 2002, where he is currently a professor in computer science.

List all the candidate context tuples that would be extracted, along with their frequency of occurrence with the seed tuples (f_S) and in the corpus (f_T).

Context words	L/M/R	f_S	f_T
"Prof."	L	?	?
Add your entries here...	?	?	?

3. (20 pts) Assume that the following pattern templates are used by the AutoSlog pattern generator:

Pattern Templates
<subject> PassiveVP(verb)
<subject> ActiveVP(verb)
<subject> ActiveVP(verb) InfinitiveVP(verb)
ActiveVP <direct-object>
InfinitiveVP <direct-object>
ActiveVP(verb) InfinitiveVP(verb) <direct-object>
PassiveVP(verb) Preposition <np>
ActiveVP(verb) Preposition <np>
InfinitiveVP(verb) Preposition <np>
Noun Preposition <np>

Consider the following text:

One pig built a house of straw while the second pig built his house with sticks. They built their houses very quickly and then sang and danced all day because they were lazy. The third little pig worked hard all day and built his house with bricks. A big bad wolf saw the two little pigs while they danced and played.

Show all of the patterns that would be generated by AutoSlog-TS when it applies the pattern templates to all of the sentences above. For each pattern p_i , also track its frequency f based on the above sentences.

Pattern (p_i)	Frequency (f)
-----------------------------------	-----------------------------------

4. (20 pts) Answer the questions below based on the Basilisk algorithm for semantic class induction, using the seed words for three semantic categories (COLOR, FRUIT, and FLOWER) and pattern data shown below. The table of pattern data includes three patterns and the nouns that each pattern extracted in an imaginary corpus. For logarithms, use log base 2.

Color Seeds: white, orange, amethyst, lavender, pink

Fruit Seeds: pear, strawberry, banana, melon, peach, apple

Flower Seeds: lily, pansy, rose, tulip

Pattern	Extracted Nouns
patternA	shrub, lily, lemon, gold, amethyst, lilac, coral, orange, white, lavender
patternB	peach, lilac, maize, lemon, tulip, pear, strawberry, lily, pink
patternC	lavender, white, pansy, banana, lemon, orange, pear

- (a) Compute $R\log F(\text{patternA})$ for the COLOR category.
- (b) Compute $R\log F(\text{patternA})$ for the FRUIT category.
- (c) Compute $R\log F(\text{patternA})$ for the FLOWER category.
- (d) Compute $R\log F(\text{patternB})$ for the COLOR category.
- (e) Compute $R\log F(\text{patternB})$ for the FRUIT category.
- (f) Compute $R\log F(\text{patternB})$ for the FLOWER category.
- (g) Compute $\text{AvgLog}(\text{"lilac"})$ for the COLOR category.
- (h) Compute $\text{AvgLog}(\text{"lilac"})$ for the FLOWER category.
- (i) Compute $\text{AvgLog}(\text{"lemon"})$ for the COLOR category.
- (j) Compute $\text{AvgLog}(\text{"lemon"})$ for the FRUIT category.

5. (20 pts) For each sentence below, correctly identify the thematic role for every noun phrase in the sentence.

(a) Jonah shipped the package to Jay on Friday.

(b) The article was written by Juliet Watson.

(c) The movie gave Ron nightmares for a week.

(d) Mel Cooper, along with Lincoln Rhyme, studied the crime scene evidence with the microscope.

(e) Jan presented a new witness to the judges.

6. (22 pts) Consider the following context vectors:

$word1 : \langle 8 \ 0 \ 1 \ 4 \ 7 \ 5 \rangle$

$word2 : \langle 9 \ 2 \ 1 \ 4 \ 2 \ 1 \rangle$

$word3 : \langle 2 \ 4 \ 6 \ 0 \ 1 \ 0 \rangle$

$word4 : \langle 9 \ 0 \ 4 \ 6 \ 0 \ 1 \rangle$

Compute the similarity scores below using the word vectors above. **You must show your work! It is not sufficient to just give a single number as the final answer.**

- (a) Similarity($word1$, $word2$) using Manhattan Distance.
- (b) Similarity($word2$, $word3$) using Manhattan Distance.
- (c) Similarity($word1$, $word3$) using Jaccard Similarity.
- (d) Similarity($word2$, $word3$) using Jaccard Similarity.
- (e) Similarity($word1$, $word4$) using Dice Similarity.
- (f) Similarity($word3$, $word4$) using Dice Similarity.
- (g) Similarity($word2$, $word4$) using Cosine Similarity.
- (h) Similarity($word1$, $word2$) using Cosine Similarity.

NO EXTRA QUESTIONS FOR CS-6340 STUDENTS THIS TIME! :)

ELECTRONIC SUBMISSION INSTRUCTIONS

You should submit the answers to this assignment **in pdf format** on our course's CANVAS site by 11:59pm on TBD.