CS 7337 – Natural Language Processing Midterm Exam

**Instructions:** Clarity of answers is more important than length of answers. Although not required (unless indicated otherwise), feel free to use graphs, charts, visuals, et al in your answers if you feel these artifacts can help support your answers. There are no bonus points for using these artifacts.

**Due date:** See instructors note; submission should be in PDF or Word DOCX file format.

Q1. a. [5 pts] Define homonymy and polysemy and give an example of each.

* Homonymy: Relationship between words that are homonyms (different meanings but are pronounced the same or spelled the same or both)
  + Example: peer
    - Noun: one that is of equal standing with another
    - Verb: to look narrowly or curiously
* Polysemy: existence of many possible meanings for a word
  + Example: crane
    - Bird
    - Equipment

b. [5 pts] Define NLU and NLG and give an example of each.

* NLU (Natural Language Understanding)
  + Processing of input data (text) supplied by a user. Uses syntactic and semantic analysis to determine meaning
  + Example:
    - Sentiment Analysis
* NLG (Natural Language Generation)
  + Generation of output of text based on input by user. Creating human text based on input data
  + Example:
    - Document summarization

Q2. You are given the following grammar for expressions:

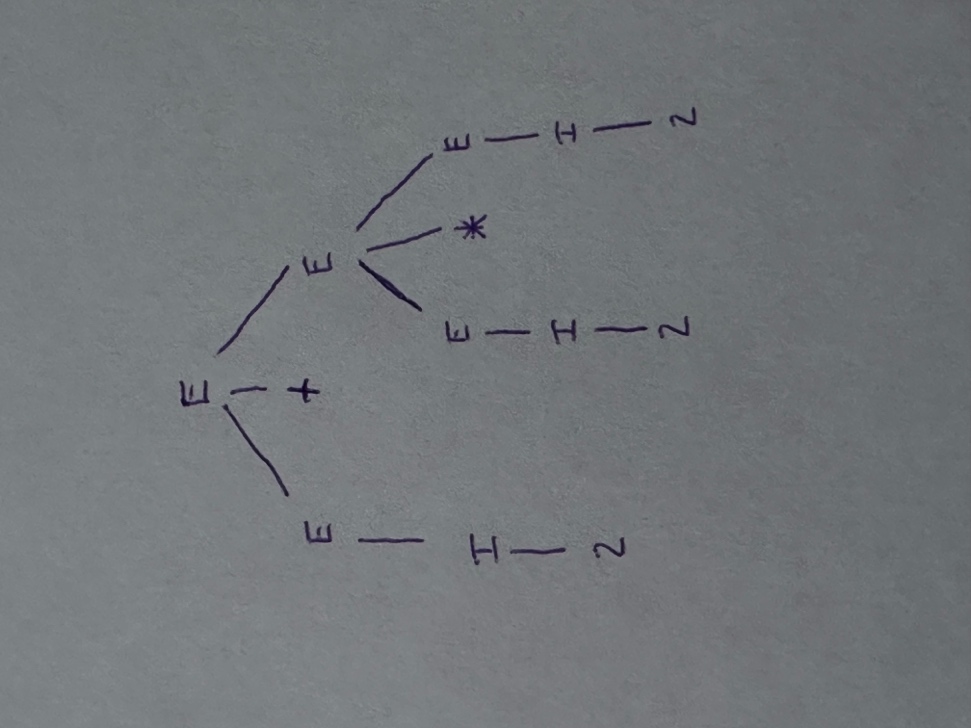
E  I I  a

E  E + E I  b

E  E \* E I  0

E  (E) I  2

1. **[10 pts] Show parse tree(s) for the expression 2 + 2 \* 2**

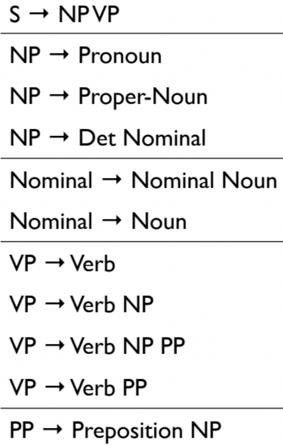
Diagram

Description automatically generated

1. **[10 pts] Describe any interesting observations in your answer to a**.

* Due to the fact we can generate two separate parse trees for a singular input, the grammar is ambiguous
* The grammar above is both left and right recursive

Q3. Consider the following grammar and sentence:



Nominal --> Nominal PP

**Sentence**: *I booked a flight from LA*

1. **[10 pts] In what way is this sentence ambiguous? Describe different interpretations of this sentence.**

* This sentence is ambiguous due to how LA is used. It could be used in either the booking or the flight. One could interpret the sentences as they booked a flight that departs from LA. Another person could interpret it as they booked a flight while they were in LA.

1. **[10 pts] Show the parse trees for this sentence and where the ambiguity manifests in the parse trees.**

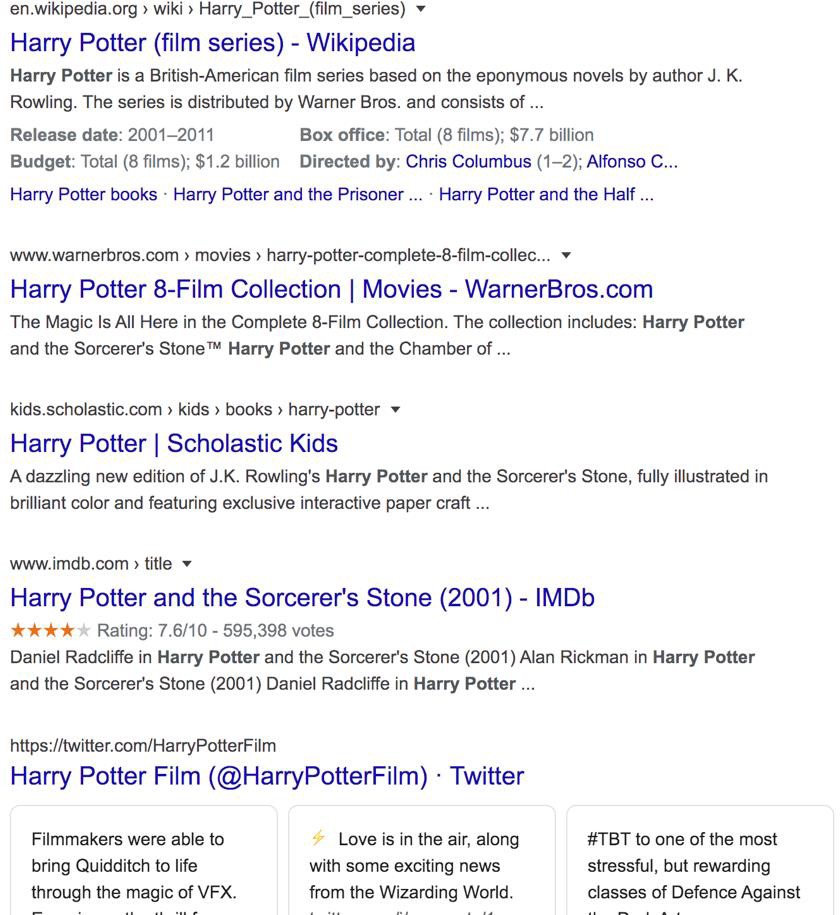
* The below parse trees show where the ambiguity manifests. The first tree on the left has the prepositional phrase as part of the verb phrase (‘booked a flight from LA’) and the second tree has it as part of a noun phrase (‘flight from LA”).

Diagram

Description automatically generatedDiagram

Description automatically generated

Q4. The image below shows Google search results for the query “harry potter”



As the results show, the query could represent any of the seven books in the harry potter franchise, any of the film adaptations of the books, a theme park, or a ride, an audiobook, cartoons, et al.

1. **[10 pts] Discuss why google shows a mix of such results and what factors can influence the search results for this query that will be presented to you.**

* Google shows a mixture of results because while there is one “Harry Potter”, it consists of seven books and eight films. Therefore, websites such as Scholastic would be returned for the book, and IMDb would be returned for the films.

1. **[15 pts] Consider the following sentence**:

*The* ***bank*** *can guarantee deposits will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.*

The word **bank** has multiple senses. Use Wordnet to show the top two sense, glossaries and examples for **bank** and describe (at a high level) how you can use this information to find the proper sense for this word in a sentence.

Wordnet link: <http://wordnetweb.princeton.edu/perl/webwn>

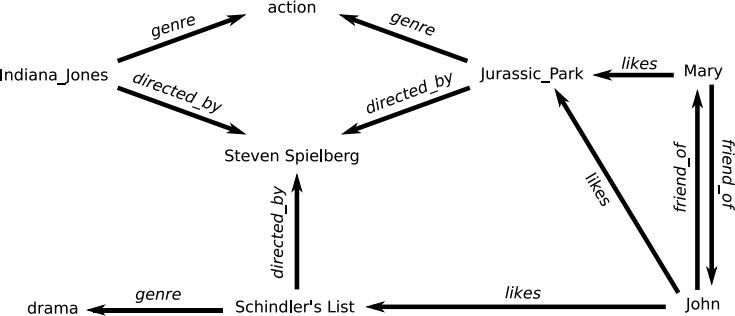
WordNet Results:

Text

Description automatically generated

* WordNet is an extremely useful took in NLP as it is a large database of semantic relations between words. Because it groups words into synsets with definitions and examples, it is much more than a thesaurus or dictionary. From the results above out of WordNet, we can see how it can be used. The key words that match within the sentence and the WordNet results are “deposits” and “mortgage” and because there is the most overlap in the second sense, the program will choose the second sense as the word.

Q5. You are building an online moving streaming service which enables looking up information on movies, genres, directors, actors and customer movie preferences.



\*\*(Indiana Jones refers to Raiders of the Lost Ark (1981))

1. **[10 pts] What is the customers intent (i.e. what are they looking for) with the following queries? (these are individual queries, not queries entered in succession)**

*“Drama”, “Jurassic Park”, “Indiana Jones: Raiders of the lost ark”, “Steven Spielberg”*

* Drama:
  + Customer is searching for films that match the genre of “drama”
* Jurassic Park
  + Customer is searching for movies titled “Jurassic Park”
* Indiana Jones: Raiders of the Lost Ark
  + Customer is searching for the exact movie “Indiana Jones: Raiders of the Lost Ark”
* Steven Spielberg
  + Customer is searching for any movie directed by Steven Spielberg

1. **[5 pts] A customer searches for “Indiana Jones” but clicks on and watches “Jurassic Park” –**

what insights can you get from this customer action?

* We can infer that the customer likes action movies that are directed by Steven Spielberg.

1. **[10 pts] The customer searches for “Indiana Jones: Raiders of the lost Ark” but it’s not available in their region (US, EU, Asia). What search results would you show the customer? Discuss how you would build that experience from a technical design perspective.**

* In the above diagram, because “Indiana Jones: Raiders of the Lost Ark” is unavailable in their region, the online service should suggest Jurassic Park as it is both an Action movie as well as directed by Steven Spielberg. In addition to the above diagram, Indiana Jones is a series of movies and other movies in the series should be included as well.
* This experience would be built out by using Semantic analysis. This would allow the streaming service to group queries from customers by utilizing input text and movies watched/clicked on to view. In addition to semantic analysis, we can utilize movie preferences to determine what movies a customer may like to watch if their preferences and past movies are similar of that to another customer or friend.