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: The relationship between words that are homonyms (words with different unrelated meanings and origins that are pronounced the same, spelled the same, or both)
  + Example:
    - peer
      * Noun: one that is of equal standing with another
    - Pier
      * A platform on pillars built from the shore into a body of water
* Polysemy: The association of a single word or phrase with multiple related meanings or variants, depending on the context in which the word or phrase is used
  + Example: foot
    - End part of leg below ankle
    - Unit of measure - root also based on the length of the body part
    - Lowest part of a mountain
    - Lower end of leg of chair or table
    - End of bed opposite from head

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 and tone
  + 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:

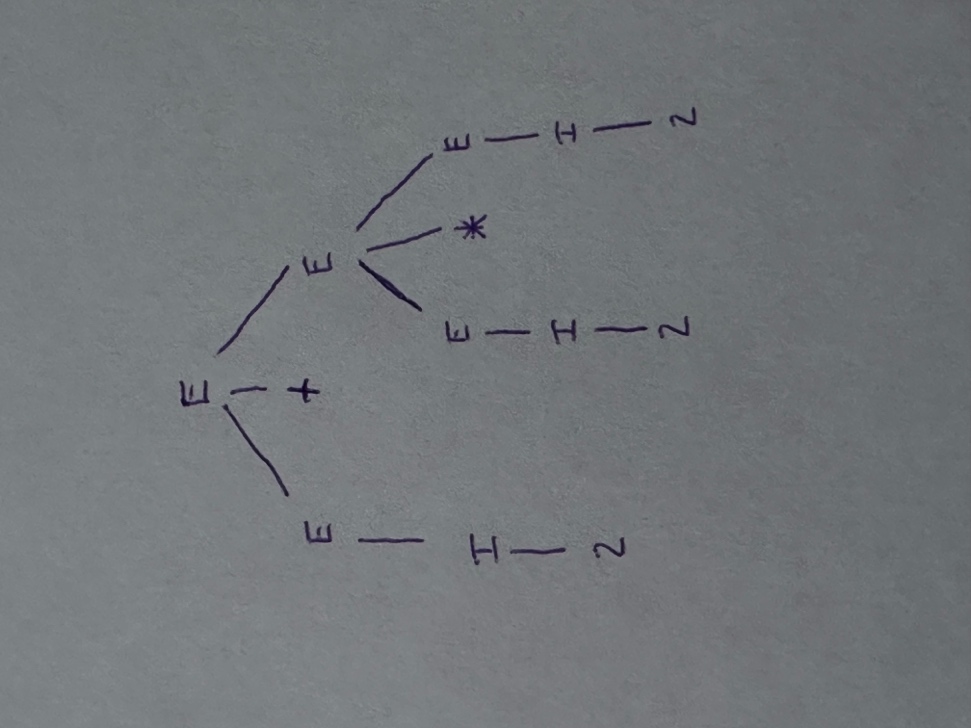
E  I I  a

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**

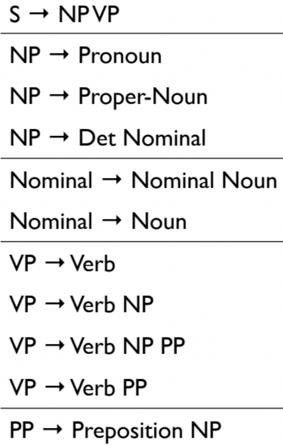
Diagram

Description automatically generated

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

* The grammar is ambiguous because two different parse trees can be made from the same input that produce the same output.
* The grammar above is both left and right recursive.
  + *Recursion* is the repeated sequential use of a particular type of linguist element or grammatical structure. Another way to describe recursion is linguistic recursion.
  + More simply, recursion has also been described as the ability to place one component inside another component of the same kind.
  + A linguistic element or grammatical structure that can be used repeatedly in a sequence is said to be *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 because of the phrase “from LA.” It could be describing the location of either the booking or the flight. One person could interpret the sentence as someone booking a flight that departed from LA. Another person could interpret it as someone booking a flight while in LA. This ambiguity changes the interpretation as to what happened in LA.

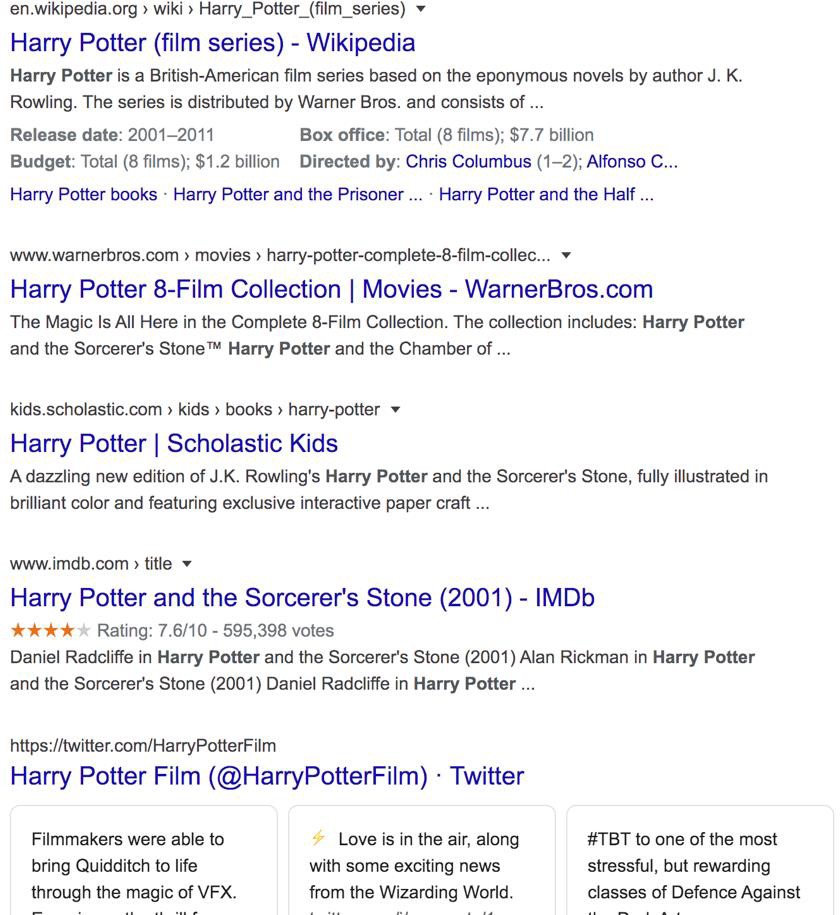
1. **[10 pts] Show the parse trees for this sentence and where the ambiguity manifests in the parse trees.**

* 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 the term “Harry Potter” is polysemantic and has multiple related meanings. It can mean any of seven books, eight films, one character, etc. The ambiguity of the term causes websites such as Scholastic to be returned for the book, and IMDb to be returned for the films. Without contextual factors, it is difficult to determine the intent of the search. However, applying semantic and syntactic analysis to the user’s search history and selection history would narrow the search to more appropriate results that match the user’s intent.

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>

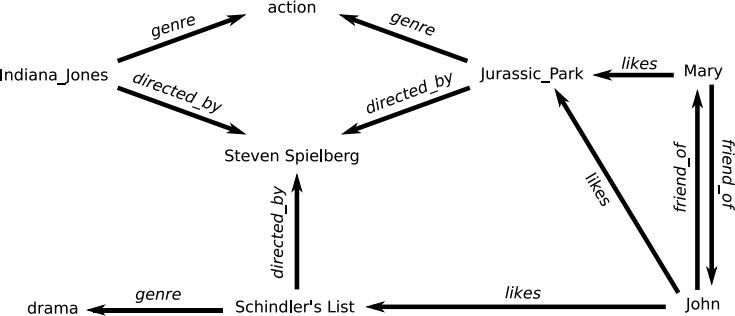
WordNet Results:

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 with titles that include the words Jurassic Park, which would include three films: *Jurassic Park, The Lost World: Jurassic Park, and Jurassic Park III.*
* 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 a combination of semantic and syntactic analysis. This would allow the streaming service to group queries from customers by utilizing input text of genres, titles, actors, and movies watched/clicked on to view. In addition to semantic analysis, the service could utilize movie preferences to determine what movies a customer may like to watch if their preferences and past movies are similar to that of another customer or friend.