

NLP Midterm

Julia Layne

Q1.

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

Homonymy is a word that has multiple senses. The meanings of these words require the context of surrounding words and corpus to get the correct meaning. Bank is often used an example to differentiate from polysemy. A bank can be a financial institute, the land near a river, or having a set of something put away. These are not tied together by a single lemma, but all have the exact same spelling.

“Bank.” *Merriam-Webster.com Dictionary*, Merriam-Webster, <https://www.merriam-webster.com/dictionary/bank>. Accessed 8 Mar. 2021.

Polysemy is a word that has multiple word senses. With these words, we must pull in context of the sentence and document to determine which meaning is the most likely. Unlike homonyms, they are all related to the same root lemma. For example `crane` can mean a species of long necked bird, a machine with a long neck, or moving one's neck to look at something. All those meanings are tied in a sense to the idea of a neck.

“Crane.” *Merriam-Webster.com Dictionary*, Merriam-Webster, <https://www.merriam-webster.com/dictionary/crane>. Accessed 8 Mar. 2021.

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

NLU is Natural Language Understanding. This is the process breaking down text and parsing out the meaning. Amazon's Alexa does this when you ask it a question. After the voice has been turned into strings, Alexa must then pull out the meaning of the tokens and then the request as a whole.

NLG is Natural Language Generation. Taking data stored in a computer and putting that into a string that looks like a natural language (preferably like another person writing/talking) with the purpose of conveying information. When an Alexa has found the correct answer to a request, it must be put into sentence form for the device to read back to a person. Putting the response data into a human readable form is natural language generation.

Q2.

You are given the following grammar for expressions:

$E \rightarrow I$	$I \rightarrow a$
$E \rightarrow E + E$	$I \rightarrow b$
$E \rightarrow E * E$	$I \rightarrow 0$
$E \rightarrow (E)$	$I \rightarrow 2$

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

Tree 1:	Tree 2:
<pre>graph TD E1[E] --- E2[E] E1 --- P1[+] E1 --- E3[E] E2 --- I1[I] I1 --- 21[2] E3 --- E4[E] E3 --- P2[*] E3 --- E5[E] E4 --- E6[E] E4 --- P3[+] E4 --- E7[E] E6 --- I2[I] I2 --- 22[2] E7 --- I3[I] I3 --- 23[2]</pre>	<pre>graph TD E1[E] --- E2[E] E1 --- P1[*] E1 --- E3[E] E2 --- E4[E] E2 --- P2[*] E2 --- E5[E] E4 --- I1[I] I1 --- 21[2] E5 --- I2[I] I2 --- 22[2] E3 --- I3[I] I3 --- 23[2]</pre>

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

Because there isn't a hierarchy between $+$ and $*$ the two trees have different orders:

Tree 1: $2 + (2 * 2) = 2 + 4 = 6$

Tree 2: $(2 + 2) * 2 = 4 * 2 = 8$

This ambiguity leads to the two separate answers for where the implied parenthesis are in the expression.

Q3. Consider the following grammar and sentence

S	→ NPVP
NP	→ Pronoun
NP	→ Proper-Noun
NP	→ Det Nominal
Nominal	→ Nominal Noun
Nominal	→ Noun
VP	→ Verb
VP	→ Verb NP
VP	→ Verb NP PP
VP	→ Verb PP
PP	→ Preposition NP
Nominal	→ Nominal PP

Sentence: I booked a flight from LA

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

The biggest ambiguity is in relation to where the flight they reserved is leaving from or where they booked it. "From LA" has two meanings:

- (Likely) Where the person booking got a flight that leaves from LA.
- (Less likely) The person booking a flight was in LA at the time that they booked it (but the flight could be leaving from anywhere).

Somewhat nonsensical versions of book in this context:

- The person brought criminal charges on the flight from LA.

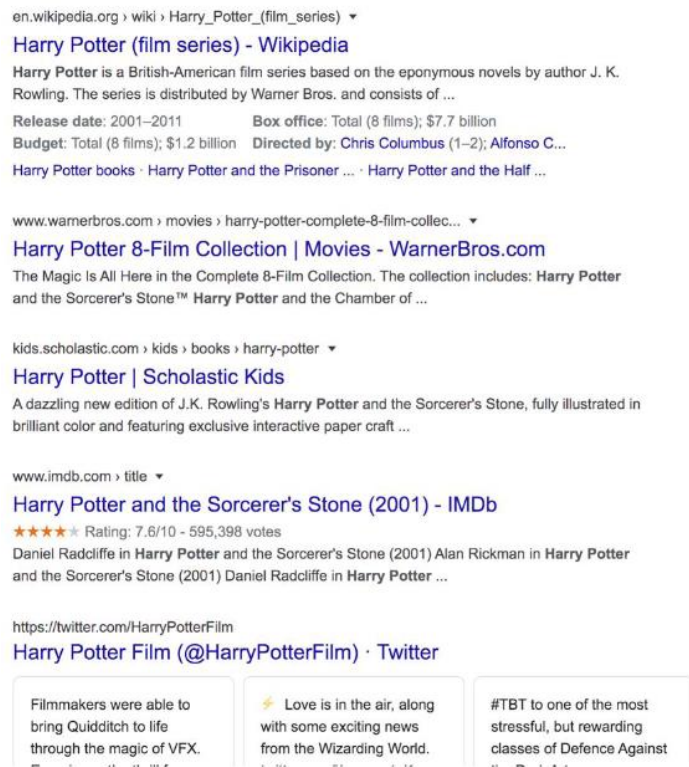
b. [10 pts] Show the parse trees for this sentence and where the ambiguity manifests in the parse trees.

Looking only at the examples where 'booked' means reserve. We see in the trees that the ambiguity is in if from LA belongs to where it was booked, or where the flight is from.

The person booking got a flight that leaves from LA. From LA is in relation to the flight.	The person booking a flight was in LA at the time that they booked it. 'From LA' is in relation to the verb 'booked'.
<pre> graph TD S --> NP S --> VP NP --> I VP --> V[booked] VP --> Nominal Nominal --> Nominal Nominal --> PP[from LA] </pre>	<pre> graph TD S --> NP S --> VP NP --> I VP --> V[booked] VP --> NP[a flight] VP --> PP[from LA] </pre>

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.

[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.

Because of the multiple types of websites associated with Harry Potter, the search algorithm will need to disambiguate through query expansion. The search query expansion will find several topics related to Harry Potter in their knowledge graph (similar to

the Tan Tan noodle example in Uber Eats Query Understanding example).

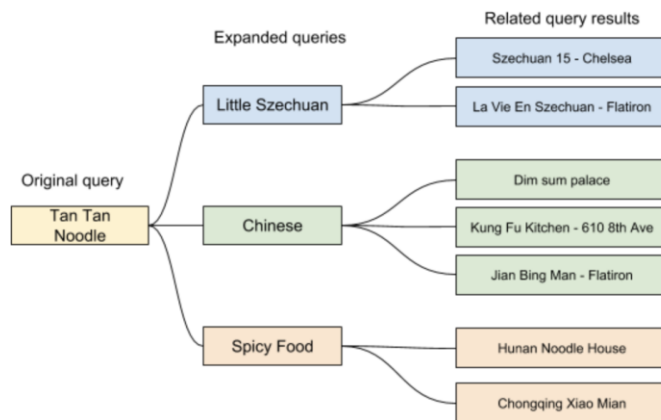


Figure 5: "Tan Tan Noodle" expands to three queries that further retrieve a set of relevant restaurant.

In this case the Original Query is 'Harry Potter'. The expanded queries found would be: books, film, theme park, ride, audiobook, etc. From those expanded topics, Google is selecting some related query results from each to give the greatest chance of answering the true intent of the user. While they may have Googled 'Harry Potter' their intent may have been to find the Harry Potter movies and they won't want to dig through all the book related links before finding the movies.

Because these results show two links to the films first, this may be because the user is in a demographic more likely to watch the movies rather than book. Or they may have a search history of looking up movies through Google. Because of this history, Google can order the links based on expected priority of the individual user in addition the overall built knowledge graphs.

[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>

First we could run the sentence through a POS tagger and see that bank is a noun. Because bank is being used as a noun, we will use the top two noun senses from Wordnet

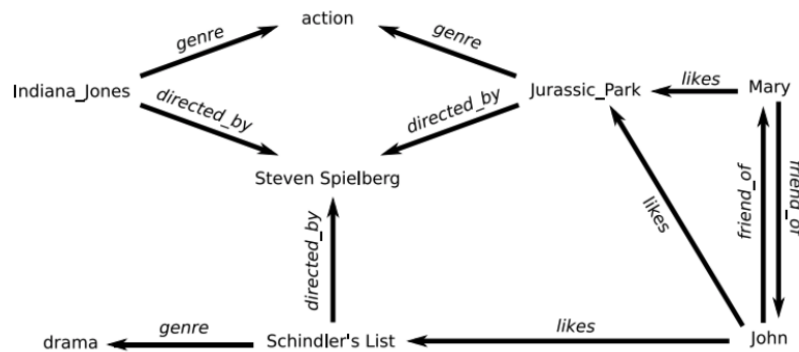
- S: (n) bank (sloping land (especially the slope beside a body of water)) "they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents"
 - o direct hyponym / full hyponym
 - o S: (n) riverbank, riverside (the bank of a river)
 - o S: (n) waterside (land bordering a body of water)
 - o direct hypernym / inherited hypernym / sister term
 - o S: (n) slope, incline, side (an elevated geological formation) "he climbed the steep slope"; "the house was built on the side of a mountain"
 - o derivationally related form
 - o W: (v) bank [Related to: bank] (enclose with a bank) "bank roads"
- S: (n) depository financial institution, bank, banking concern, banking company (a financial institution that accepts deposits and channels the money into lending activities) "he cashed a check at the bank"; "that bank holds the mortgage on my home"
 - o • direct hyponym / full hyponym
 - o S: (n) credit union (a cooperative depository financial institution whose members can obtain loans from their combined savings)
 - o S: (n) Federal Reserve Bank, reserve bank (one of 12 regional banks that monitor and act as depositories for banks in their region)
 - o S: (n) agent bank (a bank that acts as an agent for a foreign bank)
 - o S: (n) commercial bank, full service bank (a financial institution that accepts demand deposits and makes loans and provides other services for the public)
 - o S: (n) state bank (a bank chartered by a state rather than by the federal government)
 - o S: (n) lead bank, agent bank (a bank named by a lending syndicate of several banks to protect their interests)
 - o S: (n) member bank (a bank that is a member of the Federal Reserve System)
 - o S: (n) merchant bank, acquirer (a credit card processing bank; merchants receive credit for credit card receipts less a processing fee)

- S: (n) acquirer (a corporation gaining financial control over another corporation or financial institution through a payment in cash or an exchange of stock)
- S: (n) thrift institution (a depository financial institution intended to encourage personal savings and home buying)
- S: (n) Home Loan Bank (one of 11 regional banks that monitor and make short-term credit advances to thrift institutions in their region)
- • member holonym
- S: (n) banking industry, banking system (banks collectively)
- • direct hypernym / inherited hypernym / sister term
- S: (n) financial institution, financial organization, financial organisation (an institution (public or private) that collects funds (from the public or other institutions) and invests them in financial assets)
- • derivationally related form
- W: (v) bank [Related to: bank] (do business with a bank or keep an account at a bank) "Where do you bank in this town?"
- W: (v) bank [Related to: bank] (be in the banking business)
- W: (v) bank [Related to: bank] (put into a bank account) "She deposits her paycheck every month"

We could use the context from the example sentences to see related words to each version. The first bank has words related to water: canoe, river, currents. The second bank has terms related to finance: depository, deposit, money, lending, cash, check, mortgage.

Based on these relations in word net, we could look at the context in this example sentence by checking shared words between our sentence and the expanded glossary of the word sense. The second senses extended glossary shares the words: deposits, invests, mortgage. All these are also related to finance. We would expect the second wordnet sense to be the one applicable to our given sentence.

Q5. You are building an online movie 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))

[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)

Query	Intent
"Drama"	Finding a movie in the drama Genre, in this case it would be Schindler's List (in this diagram)
"Jurassic Park"	Find the specific movie Jurassic_Park (in this diagram)
"Indiana Jones: Raiders of the lost ark"	Find the specific movie Indiana_Jones (in this diagram)
"Steven Spielberg"	Finding a movie by the director Steven Spielberg. In this case a list of movies: Jurassic_Park Indiana_Jones Schindler's List

[5 pts] A customer searches for “Indiana Jones” but clicks on and watches “Jurassic Park” –what insights can you get from this customer action?

Jurassic Park is a movie that is similar to Indiana Jones. We can see that they are close in our relationship schema, but this lets us know that customers consider them similar movies.

[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.

We should show movies similar to the movie searched. Jurassic Park would likely be the first listed because it is similar in Genre as well as director. Also, if the person searching were Mary or John, both like Jurassic Park. This should be weighted into the response, if the person querying has already rated Jurassic Park and did not like it, the next most related should be shown first even though it only has the director in common.

This could be accomplished by traversing the tree and initially creating a list of other movies based on the distance of tree traversal and number of links

- Jurassic Park is linked with a distance of 1 hop in between for Director
- Schindler’s List is linked with a distance of 1 hop in between for Director
- Jurassic Park is linked with a distance of 1 hop in between for genre

Since Jurassic Park has the same distance as Schindler’s list, we might weight Jurassic Park heavier for every additional linking it has to the searched movie.

We would return [Jurassic Park, Schindler’s List]

On the additional side of preferences:

We have the data that the user has movies they have previously liked. Because of this data, the output may change based on who is doing the search.

Mary likes Jurassic Park. If we knew people are less likely to watch something they have already rated, we that might switch the order to: [Schindler’s List, Jurassic Park]
We want to show a new movie to Mary that she is likely to watch.

Since John already likes Jurassic Park and Schindler’s List too, we’d show him the original ordering.