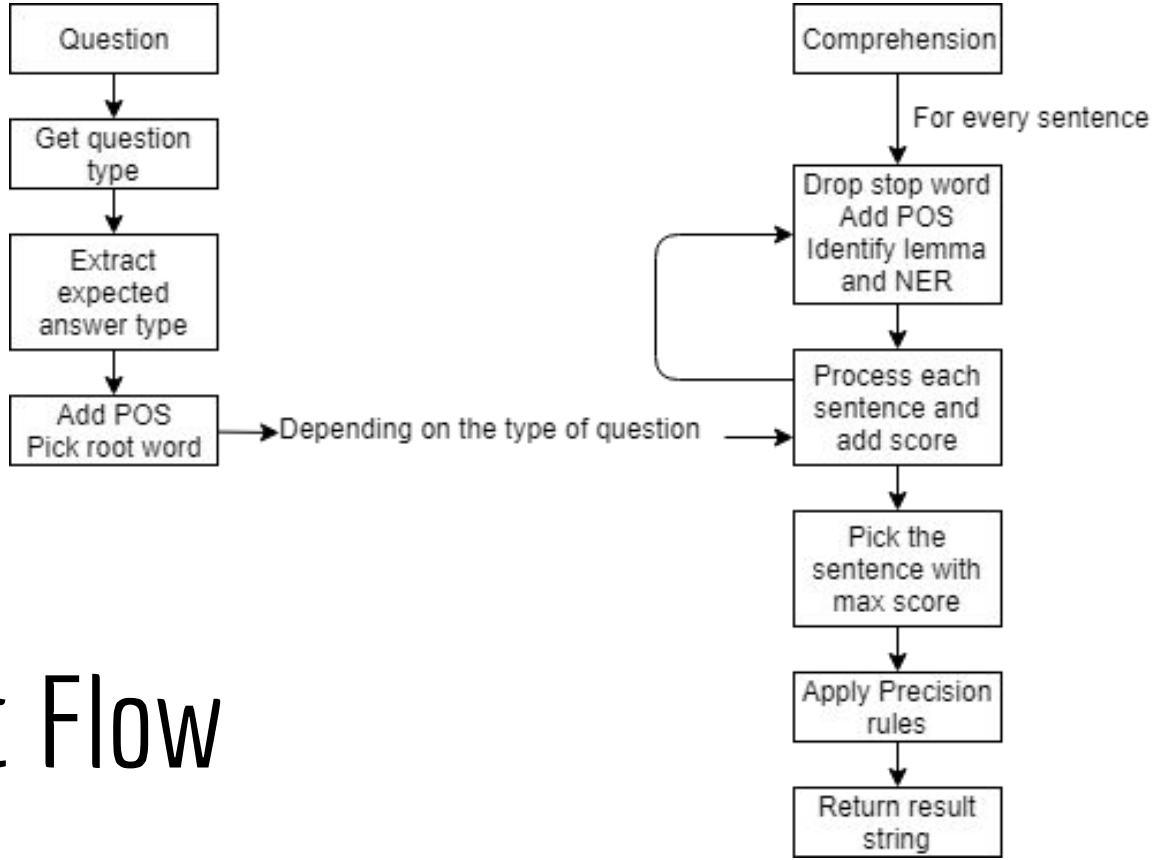
THE UNIVERSITY OF UTAH

Word Mavericks

Natural Language Processing
CS 5340 Fall 2018
Instructor: Ellen Riloff

By
Antara Bahursettiwar
Neha Kherde

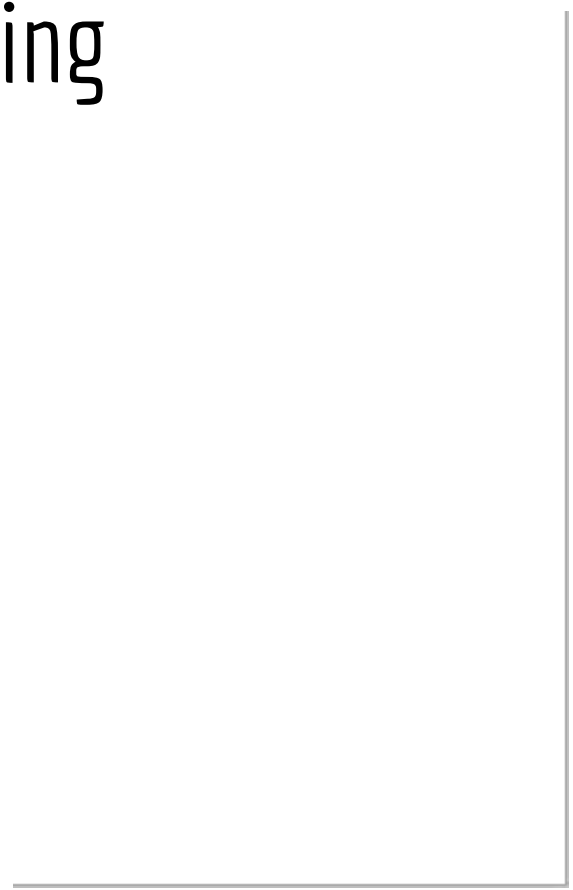
Basic Flow


Question
Get question type
Extract expected answer type
Add POS
Pick root word
Depending on the type of question
Comprehension
For every sentence
Drop stop word
Add POS
Identify lemma and NER
Process each sentence and add score
Pick the sentence with max score
Apply Precision rules
Return result string

Question Preprocessing

THE UNIVERSITY OF UTAH

Story Preprocessing



System Design & Components

THE UNIVERSITY OF UTAH

Sentence scoring on word match /similarity

- Term frequency calculation.
- For verbs, check it's similarity using **wordnet** word similarity based on **Wu Palmer's** measure.
- Score calculation based on matching words and similar verbs.

Finding the right answer..

THE UNIVERSITY OF UTAH

What?

- ➔ **Name:** What is his name? If **‘name’** in question, check for **name/call/known** in answer.
- ➔ **Time:** if **DATE** in question? Check for **today/tomorrow/yesterday** in answer.
- Words Overlap score(based on TF) + **VERB** Matches.
- Additional score if **ROOT verb** in question matched.
- Extract the sentence starting from the verb matched.

Finding the right answer..

THE UNIVERSITY OF UTAH

Why?

- Words Overlap score(based on TF) + **ADJECTIVE** matches.
- Reasoning based, so focus on '**so', 'because', 'to', 'by', 'want'**
- Check for following or preceding sentences which might have the reason.
- Extract answer starting from reason keywords

Finding the right sentence..

THE UNIVERSITY OF UTAH

When?

- Check for Named Entities **DATE** and **TIME.**
- If '**last' or 'start' or 'begin**' in question:
  Check for words:
  'last', 'first', 'since', 'ago', 'start', 'begin', 'since', 'year' in Answer.
- Extract **DATE, TIME** named entities in sentence, if found.

Finding the right sentence..

Where?

- Check for Location Named entities. **(LOC, GPE)**
- Check for prepositions: **‘in’, ‘at’, near’, ‘inside’, ‘from’** in answer.
- Extract part of sentence starting from prepositions or containing Location named entities.

Finding the right answer...

How?

Expected named entities:

- how tall, how big, how big, how high, how large, how deep: *['QUANTITY']*
- how much: *['PERCENT', 'MONEY', 'CARDINA']*
- how many: *['ORDINAL', 'QUANTITY', 'CARDINAL']*
- how long: *['QUANTITY', 'DATE', 'TIME']*
- how old, how often: *['DATE']*
- how far: *["TIME", "QUANTITY"]*

For precision: Pick Named Entity based on what the question expects

Finding the right answer...

Who?

1. Question Type: **'Who is'** Eg. Who is George Bush?
2. Question form: **'Who is' + PROPER_NOUN**
3. Check for "Who is" + ["SPACE", "PROPER NOUN", "PUNCT", "DET"]
4. Pick all "PROPER NOUN" , say name
5. Search for name in original sentence
6. Return sentence of **first match**

Precision

Only pick part of sentence that is **after the name**

Finding the right answer...

Who?

If '**PERSON**' not in Named Entity list of **Question**:

    If 'PERSON' || 'FAC' || 'name' in sentence:

        Increase score

Else if **'ORG**' not in Named Entity list of **Question**:

    If 'ORG' || 'name' in sentence:

        Increase score

Summed up all the scores from above, matching words returned max score sentence

Precision:

Returned string of Named Entity ("PERSON", "ORG", "NORP", "GPE")

Regrets & Success

THE UNIVERSITY OF UTAH

Regrets:

1. We could implement more of NLP algorithms like ngrams, BIO tagging, machine learning and coreference resolution.

Success:

1. **Term frequency** and **word** similarity gave a boost to our system.
2. **Removing** the words **common** in question and answer increased our precision.

Finding the right answer...

For **How does**, **How did, Whom, Whose**

Did Overlap and returned sentence with max score

For precision:

1. Removed overlapping words
2. In case of direct speech sentences, removed words (said, says)

Performance

Mid Point Results: 0.3600 (R: 0.4049 ,P: 0.3241)

Final Results: 0.4972 (R: 0.5686 ,P: 0.4417)

1. Mid point evaluation
2. - **Term frequency** and similarity based
   - Match **ROOT** verb in question with verb in sentence.

3.- 4 Answered multiple questions which weren't handled before
   Eg. Whom, Which
5. Removing intersection

Team Member Contributions

THE UNIVERSITY OF UTAH

Antara Bahursettiwar

Worked on Term frequency and word similarity to extract sentences. Analysed and created rules for sentence extraction and precision for Wh questions What, When, Why.

Neha Kherde

Worked on figuring out the methods to extract sentences and precision for Who, How, Where.

We pair programmed most of the code

External Resources

THE UNIVERSITY OF UTAH

Spacy

NLTK Tool Kit

WordNet Toolkit

THE UNIVERSITY OF UTAH

Thank you!

Any Questions?