NLP Project Report

● ● ●

Youce Ji, Isabel Rozario, Jingyuan Xing, Huan Zhang

Application Structure
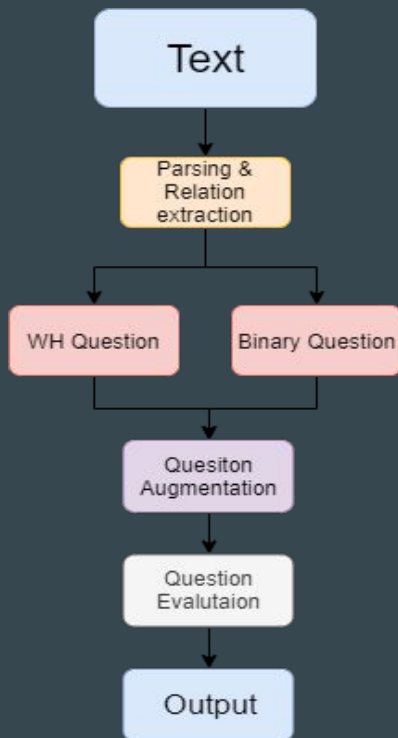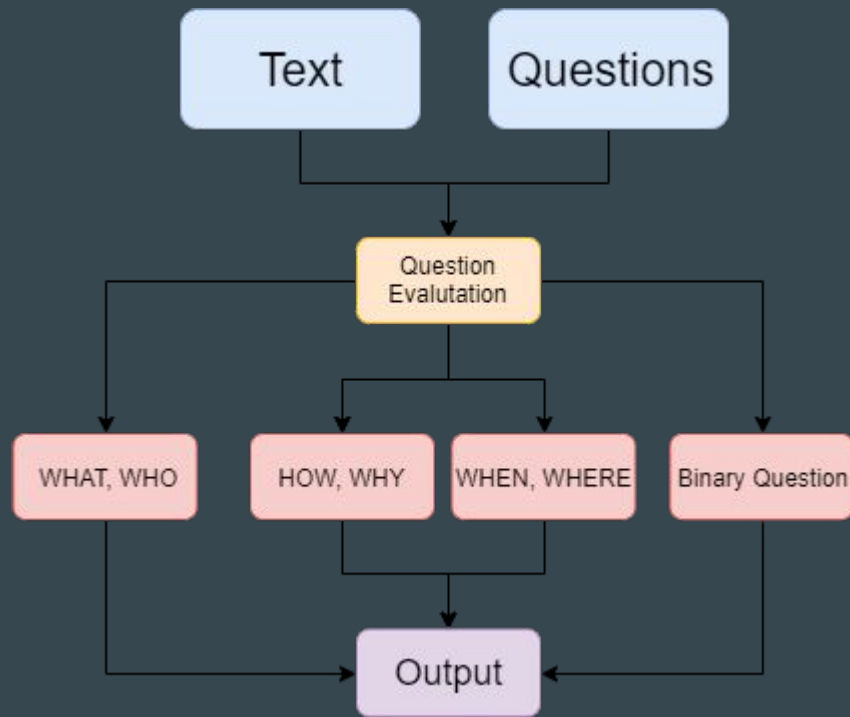
Information Extraction

Relation Extraction

- Main Idea:
  - Identify the root of sentence and extract corresponding relations into [subject, verb, object]
  - Parse every sentence based on root through dependency parser in SpaCy:
    - E.g. I have an apple → root = "have" → ['I', "have", "an apple"]
  - Other than the root, we also pay attention to other potential valid verbs/actions in each sentence
    - E.g. I have an apple, which was eaten by Jack → focus on "eaten" → ["an apple", "was eaten by", "Jack"]
  - Optional: Inclusion of context (time/place/conditions)
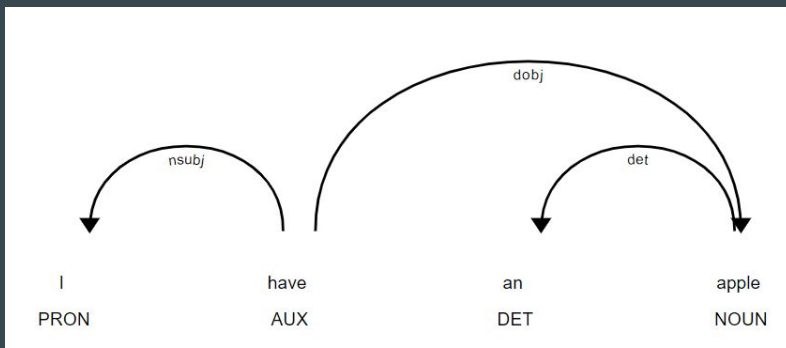
Question Asking

Wh- questions

01 **Subject-Object what question**
- (The temperature, dropped dramatically, to 0 celciums degree)
- What dropped dramatically to 0 celsius degree?

02 **Six classes of named entity**
- PERSON, LOC, DATE, EVENT, GPE, ORG
- Corresponding to When, Where, Why, etc

03 **Tense correction**
- Where is the old kingdom?
- Where was the old kingdom?
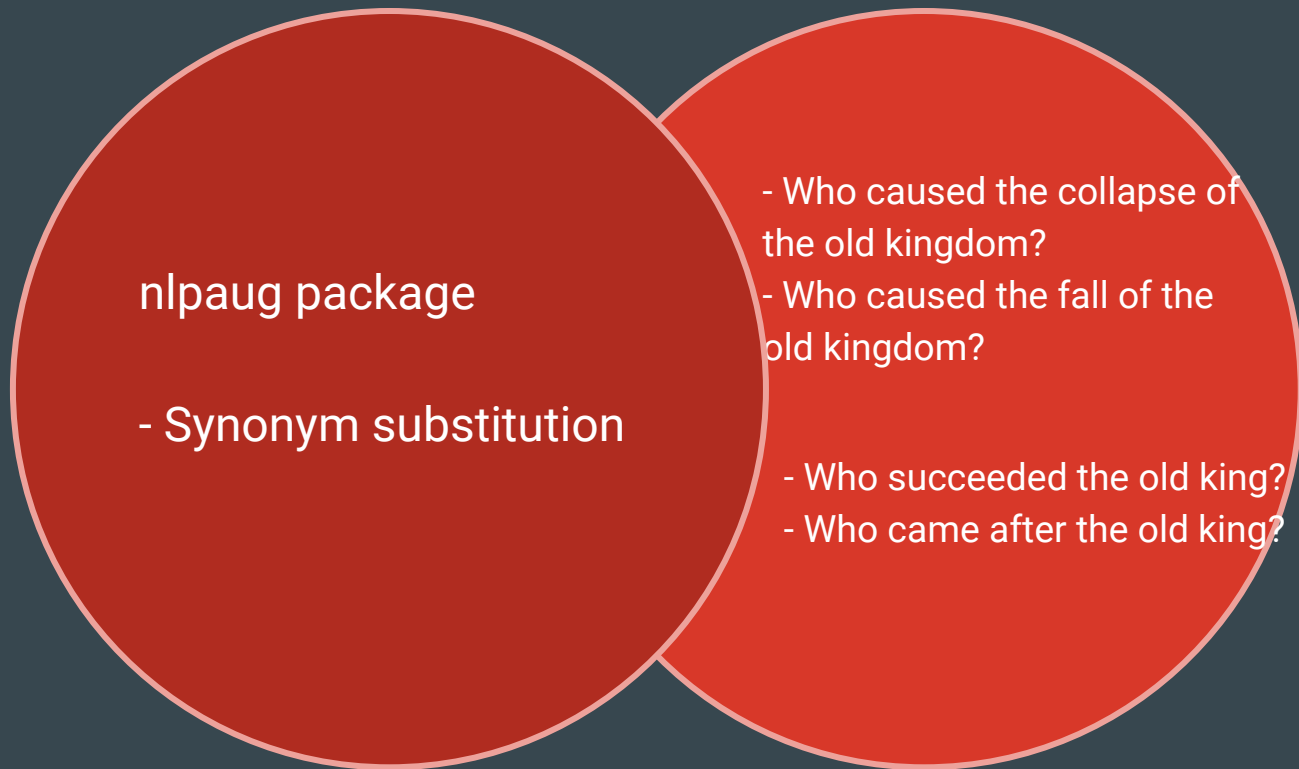
Binary questions

Auxiliary Verbs:

Is/Are/Was/Were
Has/Have
Must
…

- The first Pharaoh of the old kingdom was Djoser
- Was the first Pharaoh of the old kingdom Djoser?

- The people in Egypt must fight the intruders
- Must the people in Egypt fight the intruders
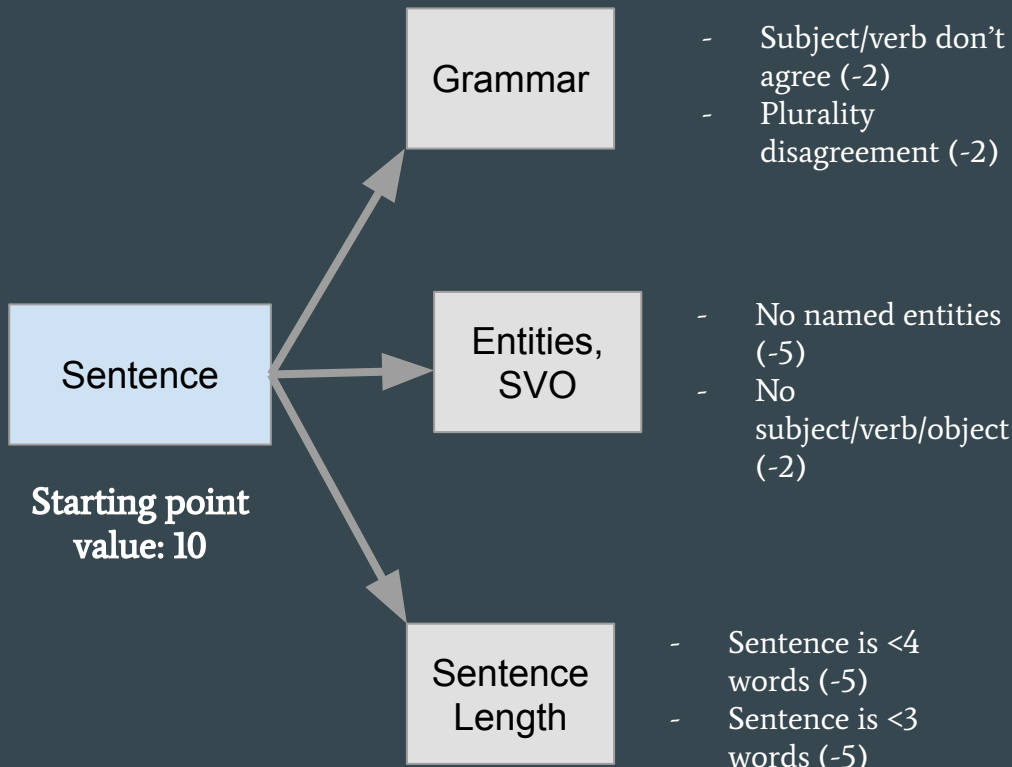
Question augmentation

nlpaug package

- Synonym substitution

- Who caused the collapse of the old kingdom?
- Who caused the fall of the old kingdom?

- Who succeeded the old king?
- Who came after the old king?

Question Answering

Sentence Ranker

- Rank the quality of a question or answer based on several characteristics
  - Helper function converts questions into declarative sentences

Sentence

Starting point value: 10

Grammar
- Subject/verb don't agree (-2)
- Plurality disagreement (-2)

Entities, SVO
- No named entities (-5)
- No subject/verb/object (-2)

Sentence Length
- Sentence is <4 words (-5)
- Sentence is <3 words (-5)

Evaluating question - sentence similarity

- For each sentence in the text, evaluate its similarity with the question.
- Assign score from 0 to 10, and the sentence with higher similarity will likely to contain the answer.
- Word matching: evaluate by how much repeated vocabulary used.
- Semantic similarity Spacy function: determine the semantic similarity of vocabs based on word vectors.

Who/What/Which Questions

Spacy Part of Speech Matching:

➤ **Subjects: NSUBJ, NSUBJPASS, CSUBJ, CSUBJPASS, AGENT, EXPL**
➤ **Objects: DOBJ, DATIVE, ATTR, OPRD**

Relation Extractions:

➤ **Subjects: Generate subject phrase by verb phrase**
➤ **Objects: Generate object phrase by verb phrase**

4 Answer Candidates

Asking for subjects or objects?
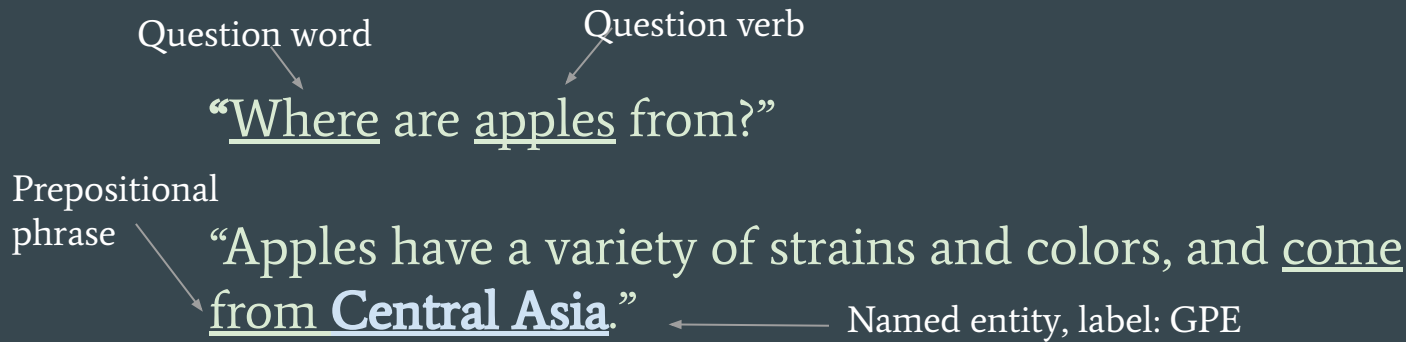
2 Answer Candidates

Evaluating these two candidates?

1 Final Answer!

When/Where Questions

- Locate named entities within the candidate sentence
  - "where" → GPE, LOC, FAC, ORG, EVENT
  - "when" → DATE, TIME, PERCENT, MONEY, QUANTITY, ORDINAL, CARDINAL
- Find the prepositional phrase containing the most relevant/greatest number of named entities

Question word

Question verb

"Where are apples from?"

Prepositional phrase

"Apples have a variety of strains and colors, and come from **Central Asia**."

Named entity, label: GPE

How/Why Question Answering

- Emphasis on prepositions and specific phrases: "through", "by", "because", "due to" and extract dependencies
- For Why questions, we try to search for reasons in nearby sentences (the one before or one after)

Binary Questions

- Comparison between Question and the Sentence identified for having the answer
- Check if every word in the question exists in the sentence or is similar enough compared.
- Dealing with double negation

Looking forward...

-   Entity coreference to remove ambiguity
-   Write more rules for different sentences structures
-   Not quite scalable for large-scale Q-A answering
    -   Alternative: Transformer Model (Attention is all you need)

Thank You !