CS 421 – Natural Language Processing – Spring 2019
Term Project (Part 2): Questions to be processed

These questions only pertain to the Movies domain. If you have to ask questions in the other domains (graduate), or you want to tackle the other domains (undergraduate extra credit), please see Section 3.

1 Sentences to be covered

You are asked to cover most of the questions described in part 1, but with some additional constraints. First we provide a general description of the questions that need to be processed; then, sec. 2.1 discusses the verb *to win*. We will include:

Yes/No questions with the verb 'to be'

(1a) Is Kubrick a director?

(1b) Is Mighty Aphrodite by Allen?

(1c) Was Loren born in Italy?

(1d) Was Birdman the best movie in 2015?

Yes/No questions with the auxiliary *Do*

(2a) Did Neeson star in Schindler's List?

(2b) Did Swank win the oscar in 2000?

(2c) Did a French actor win the oscar in 2012?

(2d) Did a movie with Neeson win the oscar for best film?

Wh-questions Crucially, we only answer Wh-questions where the *Wh-NP* is the subject, or refers to the year via *When* e.g.:

(3a) Who directed Schindler's List?

(3b) Who won the oscar for best actor in 2005?

(3c) Who directed the best movie in 2010?

(3d) Which American actress won the oscar in 2012?

(3e) Which movie won the oscar in 2000?

(3f) When did Blanchett win an oscar for best actress?

2 Comments, assumptions, simplifications etc

1. Movie titles are most often composed by more than one word. For movie titles, you can use the operator LIKE used in the example SQL query in the description of project part 2. This will also be useful for actor / director names, since the movie database provides both first and last names, but we will use only last names in the queries.

2. No determiner will be processed semantically, i.e., the semantic attachment for *a/an/the* will be empty.

3. We will exclude questions where the *WhNP* is the object, or starting with a prepositional phrase containing a *wh-* word, such as:

   (4a) Which movie did Allen direct?

   (4b) In which year did Loren win an Oscar?

   (4c) In which country was Blanchett born?

4. The only adjectives to be dealt with are *best* and nationality adjectives, such as *American, Italian, French, British, German*. Since you will have to think about semantics for each of these nationalities, you can limit yourself to these 5. You will no be required to include *supporting*, unless you wish to do so.

5. Note that the rule for S given in the example in project part 2 for *Did Allen direct Mighty Aphrodite* will not work for sentences such as *Did a French actor win the oscar in 2012?*, since the subject is not a proper noun. You will have to think how to build the semantics of an NP such as *a French actor*, and how to incorporate it into S. As already mentioned, *a / an* will not contribute anything, i.e., we will not treat the indefinite determiner as a quantifier. The "trick" of extracting the semantic variable that a formula mainly is about may be useful – see *VP.sem.variable* pg. 568 of the semantic attachment handout, here we may need to use *NP.sem.variable*.

6. We will limit ourselves to one prepositional phrase attached to an NP, e.g *a movie with Neeson*, *a movie by Kubrick*, excluding e.g. *a movie by Kubrick with Neeson*. Note that the temporal PP, e.g. *in 2010*, attaches to the VP.

2.1 The verb “to win”

The verb *to win* will require special attention. It appears in three configurations, of which we will consider two. If we classify *oscar* as PRIZE; and *movie category (best actor, actress, director etc)* as CATEGORY, we can produce the following schemata:

1. Win the oscar: schematically, `win + prize`
2. Win best actress: `win + category`
3. Win the oscar for best actress: `win + prize + category`

Our examples are all in the form (1) or (3), don’t worry about (2). However, you will have to devise appropriate semantic attachments for *win* according to these schemata, but try to be as general as possible. Note that inclusion of the category may change the answer to the question. Consider:

1. *Did Hathaway win an oscar in 2013?*: the answer is yes, she won the oscar for *best supporting actress*.
2. *Did Hathaway wi the oscar for best actress in 2013?* the answer is *no*. A real system should answer *no, but she won for best supporting actress*, but this is too complicated for us.

Note that the examples just discussed do not take into account the eventual in-PP that carries the date, as in *Did Hathaway win an oscar in 2013?*. You will have to think of a way of dealing with such PP-attachments – again the “trick” of extracting the *VP.sem.variable* may be helpful.

3 Graduate additional requirements / Undergraduate extra credit (50 points)

Extend your approach to answer questions on geography and music. For these two domains, focus on questions like (1a), (1b), (1i), (1j), (1k), (2g) from Part 1, and *yes-no* questions with the verb *to be* and containing *where*: *Where is Rome?*. Use the domain categorization module you worked on for part 1 to decide which DB to query.