A possible solution of the partial exam of INLP 2015-2016 course

Consider the following sentences got from a newspaper article:

*The price of the barrel of Western Texas oil closed to 11.56$, **its** lowest price since 1986. All the great oil producers, members of the OPEP are trying to perform agreements on production cuts in order to increase **its** price.*

1) Which are the mentions in the text corresponding to entities and which ones corresponding to events?

   Entity mentions:

   There are mentions corresponding to Named Entities, as "*Western Texas oil*" (ORG) and "*OPEP*" (ORG).
   There is also a date: "*1986*" and a money quantity: "*11.56$*".
   There are two occurrences of the pronoun "***its***".
   There are, finally, mentions corresponding to nominal phrases (NP) both simple and complex:

   The price
   The price of the barrel
   The price of the barrel of Western Texas oil
   the barrel
   the barrel of Western Texas oil
   ***its** lowest price*
   ***its** lowest price since 1986*
   the great oil producers
   All the great oil producers
   members of the OPEP
   agreements
   production cuts

   Event mentions:

   Closed
   trying
   perform

perform agreements
trying to perform agreements
increase

"Try" is just a modal that modifies "perform"
"Perform" is just a light verb (with no meaning content) so the real event is
"*trying to perform agreements*"

2) Which are the relations between these mentions?

There are textual (surface) relations between these mentions as:
"occurs before", "occurs later", "included in", "includes", etc.
For instance:

occurs_before (*"The price of the barrel of Western Texas oil", "closed"*)
included_in("*The price*", "*The price of the barrel*")
included_in("*The price of the barrel*", ("*The price of the barrel of Western Texas oil*")

Some of these relations have inverse and some are transitive.

There are, too, syntactic relations as "subject_of", "direct_object_of", "mod_of", "mod_since","mod_in_order_to","member_of"
For instance:

mod_of (*"The price","the barrel"*)
subject_of ("*The price of the barrel of Western Texas oil", "closed"*)
member_of ("*All the great oil producers*", "*OPEP*")

There are, finally, semantic relations as "agent", "theme", "experiencer", "time", "place", etc.
For instance:
experiencer (*"closed", "The price of the barrel of Western Texas oil"*)
agent ("*trying to perform agreements*", "*All the great oil producers*")

3) Some of the mentions correspond to the same entities (or events). Which ones?

"*The price of the barrel of Western Texas oil*" co-refers with the first occurrence of "***its***".
"*All the great oil producers*" and "*members of the OPEP*" also co-refer.
"*The price of the barrel of Western Texas oil*" loosely co-refers with the second occurrence of "***its***". Most of the co-reference resolution systems use to consider both mentions as co-referent although in fact the mention "*The price of the barrel of Western Texas oil*" refers to the price now while the second occurrence of "***its***" refers to the price in an undetermined date in the future.

4) Which of the relations located in 2) can be applied to the corresponding entities.

Mentions and entities are elements of different spaces. Mentions are units of the linguistic space while entities are units of the semantic space (the real world or an image of it. In Figure 1, the three mentions m_{11}, m_{12}, and m_{13}, co-refer between them and refer to the entity E_1.

Figure 1

Figure 2

5) Build a representation of the meaning of the text using a graph where nodes correspond to entities and events and edges to their relations.

A partial representation of the meaning is presented in Figure 3.

Figure 3

6) Some mentions are pronouns and some of them can co-refer with other mentions, referring together to some real world entity. Propose a simple method for co-reference resolution of the pronouns occurring in these sentences (in bold in the text).

The usual way of performing co-reference proceeds in three steps:

1. Locating co-referring mentions (pronouns in this case)
2. Locating candidate mentions, filtering out non valid ones
3. Selecting the most likely one(s).

For locating co-referring mentions we need tokenizing, morphological analysis and POS tagging. In this case, reduced to pronoun resolution, only the mentions corresponding to pronouns are chosen. From these mentions the morphological

analysis should provide useful morphological information (gender, number, animacy, type of possible co-referent, etc.).

The candidate mentions should correspond to entities (nominal). All the entity mentions are initially considered. Usually (anaphora) the co-referent occurs before the pronoun although sometimes (cataphora) the co-referent occurs after the pronoun. We propose look initially to anaphoric relations and if no candidate is selected to cataphoric ones.

The set of candidates is then filtered for discarding candidates not accomplishing the morphological constraints from the pronouns (gender, number, and the like).

The final step consists of selecting from the candidates the most likely ones. We propose the following heuristics:

1. Recency: Choose the mention candidate closest to the pronoun
2. Nesting: Choose the most external mention in the case of nested mentions.
3. Length: Prefer the longer candidates

The way of combining these heuristics can be sequential application or voting (using ranking instead of Yes/No tests).

7) **What kind of information do you need for facing the task? How this information could be obtained?**

For the first step in 6) we need a tokenizer, a morphological analyser and a POS tagger. For the second step a Name Entity Resolver, NER, is needed. Usually NER uses gazetteers (onomastics, geographics, etc.) for their task.

8) **Apply (manually) your method to solve the two reference problems in the sentences above (corresponding to the possessive pronoun *its*). Discuss the results.**

For the first step in 6) the two mentions corresponding to the occurrences of "*its*" are selected. The morphological information is the same for both mentions (non-human, singular co-referent).

For the first "*its*" the candidate set is:

The price
The price of the barrel
The price of the barrel of Western Texas oil
the barrel
the barrel of Western Texas oil
"11.56$"

No candidate from this set is removed.

From these candidates applying the heuristics the choice can be:

11.56$
The price of the barrel of Western Texas oil

For the second occurrence the process is similar, the initial candidates are:

The price
The price of the barrel
The price of the barrel of Western Texas oil
the barrel
the barrel of Western Texas oil
"11.56$"
***its** lowest price*
***its** lowest price since 1986*
OPEP
the great oil producers
All the great oil producers
members of the OPEP
agreements
production cuts

The filtering process removes all the candidates in plural (the last five).
In this case depending on the heuristics applied the result can be:

OPEP
The price of the barrel of Western Texas oil

Obviously more precise rules are needed for the 3rd step in order to select the most likely co-referent in both cases.

Barcelona, 05/11/2015