Natural Language Processing

Team Project - Part 2

Submitted By:

Anusha Sagi (ssagi2@uic.edu)

Manasa Kandimalla (mkandi3@uic.edu)

Implementation Techniques used in our project part 2?

We have used Core NLP Parser in our NLTK. We have categorized our code into 3 groups such as Movies, Music and Geography which handle questions from all these 3 categories.

Our method takes the input from the text file and generates a parse tree for the question that is asked. From this tree, we do a reverse engineering where we produce the rules that built up the tree. These are the grammar rules/production rules. Our whole process involves in using these grammar rules that were built from the tree. It verifies rules such as NNP, NN, JJ, VB etc. and maps the respective tables accordingly. The SQL query is sequentially built as these production rules are traversed. In some instances, it uses special key words like star, direct, win (in movies category), sing (in music category) and capital (in geography category) from VB -> and NN-> production rules and maps them to their respective tables. In this way, the "FROM" statements are built in the query.

The parameters in the select query are identified by two types. "Wh-questions" are identified by 'ROOT -> SBARQ' rule and "Yes/No questions" are identified by 'ROOT -> SQ' rule. After this, different rules that are generated for 'Which', 'When', 'Where' are used to identify which entities are to be selected. For 'ROOT -> SQ' rule, it is always the "select count (*)" since these are yes/no questions.

We have used rules such as NNP-> to categorize them as proper nouns. We have also used NER tagging which tags the 'PERSON', 'NATIONALITY', 'TITLE', 'CITY', 'COUNTRY' etc. in the question provided. Using all these rules and conditions, we frame the "WHERE" statements in our query.

As these rules are traversed, all the 'FROM' statements and 'WHERE' are built up. Then, all the from and where statements are merged as one at the last and finally one SQL query is built up which can be used on the database to query data.

We have a separate python file for handling queries on database. Once the query is built, it is passed to the respective database based on the category result (part 1 of the project) and the result is obtained.

This way, it traverses every question in the input file and generates answers. If, any of he input question is out of reach for our code, it answers it as "I do not know". All this is outputted to an output file which contains the question, category, parse tree, SQL query and the answer.

REPORT SAMPLE OUTPUT:

Question 1: Is Kubrick a director?

Category: Movies

Parse Tree:

(ROOT

   (SQ (VBZ Is) (NP (NNP Kubrick)) (NP (DT a) (NN director)) (. ?)))

SQL statement:

select count(*) from Person P INNER JOIN Director D on P.id = D.director_id where P.name like '%Kubrick%'

Answer:

Yes

Question 2: Is Mighty Aphrodite by Allen?

Category: Movies

Parse Tree:

(ROOT

  (SQ

    (VBZ Is)

    (NP (NNP Mighty))

    (NP (NP (NNP Aphrodite)) (PP (IN by) (NP (NNP Allen))))

    (. ?)))

SQL statement:

select count(*) from Person P INNER JOIN Director D ON P.id = D.director_id INNER JOIN Movie M ON D.movie_id = M.id where M.name like '%Mighty Aphrodite%' and P.name like '%Allen%'

Answer:

Yes

Question 3: Was Loren born in Italy?

Category: Movies

Parse Tree:

(ROOT

  (SQ

    (VBD Was)

    (NP (NNP Loren))

    (VP (VBN born) (PP (IN in) (NP (NNP Italy))))

    (. ?)))

SQL statement:

select count(*) from Person P where P.name like '%Loren%' and P.pob like '%Italy%'

Answer:

Yes

Question 4: Was Birdman the best movie in 2015?

Category: Movies

Parse Tree:

(ROOT

  (SQ

    (VBD Was)

    (NP (NNP Birdman))

(PP

  (NP (DT the) (JJS best) (NN movie))

  (PP (IN in) (NP (CD 2015))))

(. ?)))

SQL statement:

select count(*) from Movie M INNER JOIN Oscar O ON M.id = O.movie_id where M.name like '%%' and O.type like '%BEST-PICTURE%' and O.year like '%2015%'

Answer:

Yes

Question 5: Did Neeson star in Schindler's List?

Category: Movies

Parse Tree:

(ROOT

  (SQ

    (VBD Did)

    (NP (NNP Neeson))

    (VP

      (VB star)

      (PP (IN in) (NP (NP (NNP Schindler) (POS 's)) (NN List))))

    (. ?)))

SQL statement:

select count(*) from Person P INNER JOIN Actor A ON P.id = A.actor_id INNER JOIN Movie M ON A.movie_id = M.id where P.name like '%Neeson%' and M.name like '%Schindler's List%'

Answer:

Yes

Question 6: Did Swank win the oscar in 2000?

Category: Movies

Parse Tree:

(ROOT

  (SQ

    (VBD Did)

    (NP (NNP Swank))

    (VP

      (VB win)

      (NP (NP (DT the) (NN oscar)) (PP (IN in) (NP (CD 2000)))))

    (. ?)))

SQL statement:

select count(*) from Person P INNER JOIN Oscar O ON P.id = O.person_id where P.name like '%Swank%' and O.year like '%2000%'

Answer:

Yes

Question 7: Did a French actor win the oscar in 2012?

Category: Movies

Parse Tree:

(ROOT

  (SQ

    (VBD Did)

    (NP (DT a) (JJ French) (NN actor))

    (VP

      (VB win)

      (NP (NP (DT the) (NN oscar)) (PP (IN in) (NP (CD 2012)))))

    (. ?)))

SQL statement:

select count(*) from Person P INNER JOIN Oscar O ON P.id = O.person_id where P.pob like '%France%' and O.type like '%BEST-ACTOR%' and O.year like '%2012%'

Answer:

Yes

Question 8: Did a movie with Neeson win the oscar for best film?

Category: Movies

Parse Tree:

(ROOT
  (S
    (S
      (VP
        (VBD Did)
        (NP (DT a) (NN movie))
        (PP (IN with) (NP (NNP Neeson)))))
    (VP
      (VBP win)
      (NP
        (NP (DT the) (NN oscar))
        (PP (IN for) (NP (JJS best) (NN film)))))
    (. ?)))

SQL statement:

select count(*) from Person P INNER JOIN Director D ON P.id = D.director_id INNER JOIN Oscar O ON D.movie_id = O.movie_id where O.type like '%BEST-PICTURE%' and P.name like '%Neeson%'

Answer:

No

Question 9: Who directed Schindler's List?

Category: Movies

Parse Tree:

(ROOT

  (SBARQ

    (WHNP (WP Who))

    (SQ

      (VP

        (VBD directed)

        (NP (NP (NNP Schindler) (POS 's)) (NN List))))

    (. ?)))

SQL statement:

select P.name from Person P INNER JOIN Director D ON P.id = D.director_id INNER JOIN MOVIE M ON D.movie_id = M.id where M.name like '%Schindler's List%'

Answer:

Steven Spielberg

Question 10: Who won the oscar for best actor in 2005?

Category: Movies

Parse Tree:

(ROOT

  (SBARQ

    (WHNP (WP Who))

    (SQ

      (VP

        (VBD won)

        (NP (DT the) (NN oscar))

(PP (IN for) (NP (JJS best) (NN actor)))

                    (PP (IN in) (NP (CD 2005)))))

        (. ?)))

SQL statement:

select P.name from Person P INNER JOIN Oscar O ON P.id = O.person_id where O.type like '%BEST-ACTOR%' and O.year like '%2005%'

Answer:

Jamie Foxx

Question 11: Who directed the best movie in 2010?

Category: Movies

Parse Tree:

(ROOT

  (SBARQ

    (WHNP (WP Who))

    (SQ

      (VP

        (VBD directed)

        (NP

          (NP (DT the) (JJS best) (NN movie))

          (PP (IN in) (NP (CD 2010))))))

    (. ?)))

SQL statement:

select P.name from Person P INNER JOIN Director D ON P.id = D.director_id INNER JOIN OSCAR O ON D.movie_id = O.movie_id where O.type like '%BEST-PICTURE%' and O.year like '%2010%'

Answer:

Kathryn Bigelow

Question 12: Which American actress won the oscar in 2012?

Category: Movies

Parse Tree:

(ROOT

  (SBARQ

    (WHNP (WDT Which) (JJ American) (NN actress))

    (SQ

      (VP

        (VBD won)

        (NP (DT the) (NN oscar))

        (PP (IN in) (NP (CD 2012)))))

    (. ?)))

SQL statement:

select P.name from Person P INNER JOIN Oscar O ON P.id = O.person_id where P.pob like '%USA%' and O.type like '%BEST-ACTRESS%' and O.year like '%2012%'

Answer:

Meryl Streep

Question 13: Which movie won the oscar in 2000?

Category: Movies

Parse Tree:

(ROOT

  (SBARQ

    (WHNP (WDT Which) (NN movie))

    (SQ

      (VP

        (VBD won)

          (NP (DT the) (NN oscar))

          (PP (IN in) (NP (CD 2000)))))))

    (. ?)))

SQL statement:

select M.name from Movie M INNER JOIN Oscar O ON M.id = O.movie_id where O.type like '%BEST-PICTURE%' and O.year like '%2000%'

Answer:

American Beauty

Question 14: When did Blanchett win an oscar for best actress?

Category: Movies

Parse Tree:

(ROOT

  (SBARQ

    (WHADVP (WRB When))

    (SQ

      (VBD did)

      (NP (NNP Blanchett))

      (VP

        (VB win)

        (NP

          (NP (DT an) (NN oscar))

          (PP (IN for) (NP (JJS best) (NN actress))))))

    (. ?)))

SQL statement:

select O.year from Oscar O INNER JOIN Person P ON O.person_id = P.id where P.name like '%Blanchett%' and O.type like '%BEST-ACTRESS%'

Answer:

2014

Question 15: Did Madonna sing Papa Do Not Preach?

Category: Music

Parse Tree:

(ROOT

  (SQ

    (VBD Did)

    (NP (NNP Madonna))

    (VP

      (VB sing)

      (SBAR

        (S (NP (NN Papa)) (VP (VBP Do) (RB Not) (VP (VB Preach))))))

    (. ?)))

SQL statement:

select count(*) from Artist A INNER JOIN Album AL ON A.id = AL.artsitID INNER JOIN Track T ON AL.albumID = T.albumID where A.name like '%Madonna%' and T.name like '%Papa Do not Preach%'

Answer:

Yes

Question 16: Does the album Thriller include the track Beat It?

Category: Music

Parse Tree:

(ROOT

  (S

    (VP

      (VBZ Does)

```
      (SBAR

        (S

          (NP (DT the) (NN album) (NN Thriller))

          (VP

            (VBP include)

            (S (NP (DT the) (NN track) (NN Beat)) (NP (PRP It)))))))))

    (. ?)))
```

SQL statement:

select count(*) from Album AL INNER JOIN Track T ON AL.albumID = T.albumID where AL.name like '%Thriller%' and T.name like '%Beat It%'

Answer:

Yes

Question 17: Was Beyonce' born in the USA?

Category: Movies

Parse Tree:

```
(ROOT

  (S

    (VP

      (VBD Was)

      (SBAR

        (S

          (NP (NNP Beyonce) (POS '))

          (VP (VBN born) (PP (IN in) (NP (DT the) (NNP USA)))))))

    (. ?)))
```

SQL statement:

select count(*) from Person P where P.name like '%Beyonce%' and P.pob like '%USA%'

Answer:

Yes

Question 18: Which pop artist sings Crazy In Love?

Category: Music

Parse Tree:

(ROOT

  (SBARQ

    (WHNP (WDT Which) (NN pop) (NN artist))

    (SQ

      (VP

        (VBZ sings)

        (S (NP (NNP Crazy)) (PP (IN In) (NP (NNP Love))))))

    (. ?)))

SQL statement:

select A.name from Artist A INNER JOIN Album AL ON A.id = AL.artsitID INNER JOIN Track T ON AL.albumID = T.albumID where T.name like '%Crazy in Love%'

Answer:

Beyonce

Question 19: Which album by Swift was released in 2014?

Category: Music

Parse Tree:

(ROOT

  (SBARQ

    (WHNP

      (WHNP (WDT Which) (NN album))

      (PP (IN by) (NP (NNP Swift))))

(SQ (VBD was) (VP (VBN released) (PP (IN in) (NP (CD 2014)))))

(. ?)))

SQL statement:

select AL.name from Album AL INNER JOIN Artist A ON AL.artsitID = A.id where A.name like '%Swift%' and AL.releaseDate like '%2014%'

Answer:

Question 20: Where was Gaga born?

Category: Movies

Parse Tree:

(ROOT

  (SBARQ

    (WHADVP (WRB Where))

    (SQ (VBD was) (NP (NNP Gaga)) (VP (VBN born)))

    (. ?)))

SQL statement:

  from where

Answer:

I do not know

Question 21: Is Rome the capital of Italy?

Category: Geography

Parse Tree:

(ROOT

  (SQ

    (VBZ Is)

    (NP (NNP Rome))

(NP (NP (DT the) (NN capital)) (PP (IN of) (NP (NNP Italy))))

(. ?)))

SQL statement:

select count(*) from CITIES CI INNER JOIN Capitals CA on CI.Id = CA.CityId INNER JOIN Countries CY on CA.CountryId = CY.Id where CI.name like '%Rome%' and CY.name like '%Italy%'

Answer:

Yes

Question 22: Is France in Europe?

Category: Geography

Parse Tree:

(ROOT

  (SQ

    (VBZ Is)

    (NP (NNP France))

    (ADVP (IN in))

    (NP (NNP Europe))

    (. ?)))

SQL statement:

select count(*) from Countries CY INNER JOIN CountryContinents CC on CY.Id = CC.CountryId INNER JOIN Continents CO on CC.ContinentId = CO.Id where CY.name like '%France%' and CO.continent like '%Europe%'

Answer:

Yes

Question 23: Is Rome in Italy?

Category: Geography

Parse Tree:

(ROOT

   (SQ (VBZ Is) (NP (NNP Rome)) (ADVP (IN in)) (NP (NNP Italy)) (. ?)))

SQL statement:

select count(*) from CITIES CI INNER JOIN Capitals CA on CI.Id = CA.cityId INNER JOIN Countries CY on CA.CountryId = CY.Id where CI.name like '%Rome%' and CY.name like '%Italy%'

Answer:

Yes

Question 24: What is the capital of France?

Category: Geography

Parse Tree:

(ROOT

  (SBARQ

    (WHNP (WP What))

    (SQ

      (VBZ is)

      (NP (NP (DT the) (NN capital)) (PP (IN of) (NP (NNP France)))))

    (. ?)))

SQL statement:

select CI.name from CITIES CI INNER JOIN Capitals CA on CI.Id = CA.CityId INNER JOIN Countries CY on CA.CountryId = CY.Id where CY.name like '%France%'

Answer:

Paris

Question 25: Where is Rome?

Category: Geography

Parse Tree:

(ROOT

   (SBARQ (WHADVP (WRB Where)) (SQ (VBZ is) (NP (NNP Rome))) (. ?)))

SQL statement:

select CY.name from Countries CY INNER JOIN Capitals CA on CY.Id = CA.CountryId INNER JOIN Cities CI on CA.CityId = CI.Id where CI.name like '%Rome%'

Answer:

Italy

Question 26: Where is France?

Category: Geography

Parse Tree:

(ROOT

   (SBARQ (WHADVP (WRB Where)) (SQ (VBZ is) (NP (NNP France))) (. ?)))

SQL statement:

select CO.continent from Continents CO INNER JOIN CountryContinents CC on CO.Id = CC.ContinentId INNER JOIN Countries CY on CC.CountryId = CY.Id where CY.name like '%France%'

Answer:

Europe