

# CS 421 – Natural Language Processing – Spring 2016

## Term Project: Building Mini-Watson (Part 2)

Name – Surbhi Arora, UIN-660042750

Name – Shaika Chowdhury, UIN-664920919

**Objective:** Derives semantic representation from parse tree in SQL query format and returns answer of this query from DB.

### Parse Tree Traversal:

The syntactic tree generated from the parser is traversed in a top-down fashion starting from the root. In every iteration when the total number of iterations is equal to the number of children of the root - it checks if the node is of type string. If it is of type string then that means that it has reached the leaves of the tree and can return. Otherwise, it recursively calls the method with the non-terminal node to traverse that subtree. In our script it's done in GrammarRulesAndSemAttach method.

### Semantic Attachments and Lambda Reduction:

First the POS tag of the root node of the tree is checked to set the initial value of the SELECT data structure, which in turn serves as its semantic attachment. A POS tag of 'SQ'/'VBZ'/'VBD'/'VP' sets the SELECT data structure to "count\*" for yes/no question and 'WHNP'/'WHADVP' sets it to "P.NAME" for wh questions by default which will be replace to "O.YEAR" if first word is tagged as 'WRB' or "M.NAME" according to verb and noun checks.

On traversing the tree once the leaves are reached, it checks if they are proper nouns. For a proper noun the semantic attachment is simply the proper noun itself. In our script, it's done by calling the ProperNoun.sem method. Additionally, for the proper nouns to be available **for lambda reduction**, all the proper nouns of a sentence are stored in a list, namely, PropNoun\_list. The semantic attachment of the verb is made by calling the Verb.sem method and Verb.sem\_wh for yes/no and wh questions respectively. Within that method, lambda reduction is done to bound the variables of the WHERE part of the sql to the proper nouns by checking a set of conditions. For example, if the query is "Did Allen direct Mighty Aphrodite?", the proper noun list would append 'Allen', 'Mighty' and 'Aphrodite'. Here the P.NAME and M.NAME variables of the WHERE clause, "P.NAME like %Allen and M.NAME like %Mighty Aphrodite%" gets bounded to 'Allen' and 'Mighty Aphrodite' respectively by querying the database to check if 'Allen' is a person or a movie to ensure that the NAME variable of person table P is assigned 'Allen' and that of movie table M is assigned 'Mighty Aphrodite'. The FROM clauses are also concatenated with appropriate clauses.

Some query specific issues we had to deal with were:

- *year*: If a sentence has year in it, then it gets POS tagged as 'CD' and length of that word is four. The semantic attachment for this is simply the year itself which is stored in a variable during tree traversal and passed as parameters to the Verb. sem and Verb. sem\_wh methods for lambda reduction.
- *French/British/Italian/American/German*: Because certain nouns like actor/actress can be prepended by adjective like these, such adjectives are semantically attached by storing the adjective in a variable and mapping it to the corresponding country by indexing to a dictionary, namely, pos\_dict as the lambda reduction in WHERE clause requires the value as country rather than as its adjective form.
- *by*: As by indicates 'direct', it is semantically attached to director in WHERE clause.
- *best/Oscar*: As the mention of 'best' in a sentence indicates Oscar, it is semantically attached to Oscar in WHERE clause.
- Movie name with "-" or ":" : if a string name has "-" or ":" then that string will be checked in the database, to verify whether it's a movie name or not
- Concatenated movie or track name : If a string is of format uppercase (A-Z) then one or more than one lowercase (a-z) character and again uppercase (A-Z) character, it will be split with space in it and then checked in database, to verify whether it's a valid movie name or not
- To verify version of the movie name: check for a numbers is applied. If a string has a number of length less than four than it will be added in the lambda function to check in where clause.

### **SQL Query Creation:**

Once all the proper nouns have been bound to appropriate variables in WHERE clause and FROM is also set with the corresponding clause, the START, FROM and WHERE are concatenated to form the SQL query. The database is then queried with this SQL in the NewQuery method.