CS 421 – Natural Language Processing – Spring 2019
Term Project: A mini-Alexa/Siri that doesn't talk
Introduction

Introduction

Inspired by personal assistants such as Siri and Alexa, we will build a (very!) scaled down version of such a question/answering (Q/A) system, employing only text, not speech. Essentially Q/A systems retrieve information from a number of repositories of information, including databases. While Alexa and Siri rely mostly on information you can directly find on the web, we will retrieve answers from databases.

We will deal with three categories: world geography, music, and movies. You are provided with three toy databases on these three domains. You will have to parse the query, and translate it into SQL (no previous knowledge of SQL is necessary). You will have to infer the domain of the query from the query itself, so that you can access the right database. At the end of the project, we should be able to retrieve answers to questions such as *When did Streep win an oscar?* and *Who directed Silence of the Lamb?*

Due dates are as follows, all are meant as 11:59pm. Part 1 is easier than Part 2, that's why an early due date and fewer points.

	Due	Points
Declare groups	4/1 (Mon)	N/A
Part 1	4/15 (Mon)	100
Part 2	5/5 (Sun)	220

The project will be in two parts. The total score for the project is 320 points.

Important Notes

- In the first and easier part, you will choose one parser, and parse all the sentences listed in the Project Part 1 description (this folder). You will also develop a module that infers the category of the question.
- In the second and more challenging part, you will derive a semantic representation from the syntactic trees the parser returns to you. We will use SQLite and translate the language query into an SQL query, so that you will be able to compute the answer from those toy databases. See an example of how a question would be translated in the sports domain at the end of this document.
- The project should be developed in Java or Python. If your group wants to use any other programming language, you must request approval from your TA or instructor.

An example of an SQL query

To give you a sense of the semantics we're looking for, here's an example SQL query in a different domain, winter olympics: *Who arrived first in ski jumping?*

```
SELECT results.winner
FROM results, competitions
WHERE results.comp_id = competitions.comp_id
  AND medal = 'gold'
  AND name = 'skijumping';
```

We will use SQLite as our database engine. This software is freeware, easy to install, and easy to use. More details will be forthcoming, but if you want to start playing with it, and the DBs we have assembled (provided in the same folder), you can download it from the website: Sqlite download

It can be integrated with many different programming languages, or invoked from the command line. Alternatively, SQLite Manager is a Firefox / Chrome add-on: Sqlite add-on