tinyWatson Final Project

By Ishta Bhagat and Kevin Liu

The tinyWatson is a project that was part of a course for CS421 Natural Language Processing at University of Illinois. We wanted to create our own question/answer system. We used software and libraries: SQLite, NLTK, the Stanford Packages, Git Browser, and Bootstrap. We are proud of our work and are happy that we could even create something like our own miniature Watson. We integrated three databases: movies, geography, and music. So along with that, we had a great user interface design, but we could not integrate it as we wished. However, we have included it in the directory for grading purposes. In terms of the actual parsing and tagging, we did NER tagging. We stripped the string getting rid of special characters, and tokenizing it. We then would parse based on what type of question it was like. An example would be the WH- type questions or the yes/no answered questions. We also sorted for if a person was mentioned or a location. We had a condition for time and whether it was in scope for our data. We created parse trees for each string input and tagged each token. Thus, after categorizing the words into their respective categories, we would call a sql query to retrieve the information from the database. We did not know how to query and work with SQLite so that was a new experience for us. In term of challenges, we did not have much time to work on it, but we accomplished it. We wanted to do a second option where the user could answer a question we/the system asked and we would parse their answer to see if they were correct. Unfortunately, due to time, we could not complete it. We had a UI set up for it as well, but we could not finish the parsing on time so we did not include it in the files. We, also, had a great user interface design, but due to many issues involving python, apache, simple server and over browser

restrictions, we redid a simpler user interface in python itself. Despite all this, we got all of it to

work.