



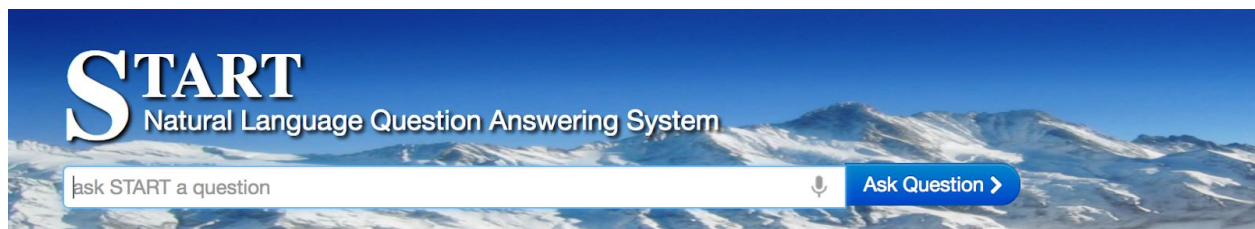
CS/NEUR125 Brains, Minds, and Machines

Lab 7: Exploring Online Language Understanding Systems

Due: Wednesday, March 22

This lab is an open-ended exploration of some interactive online computer systems that embody natural language understanding. These systems include the [START Natural Language Question Answering System](#) that can answer questions on a wide range of topics, [Google Translate](#) that translates text between two languages, and the [Mitsuku Chatbot](#) that converses with the user. You will also consider the question-answering ability of [IBM's Watson](#) that has beaten top human players at the game of Jeopardy. As usual, you will work with a partner, but we encourage all of you to work together on this lab, to share ideas about things to try and thoughts on what you learned from your exploration, about the challenges, limitations (and successes!) of these systems.

To begin, create a copy of this Google document, modify the title of the copy to include your partner names, and share the copy between partners, as you did in previous labs. The questions for you to answer are [shown in blue](#).



[START](#) is the first Web-based question answering system to be open to the world, and has been operating online continuously since December, 1993. It was developed by Boris Katz and his colleagues at MIT and is continually enhanced with new knowledge and language understanding abilities. Some sample questions about geography, science, arts and entertainment, and history and culture, can be found at the home page linked above, and a brief description of how the system works can be found [here](#).

Q1. Visit [START](#) and try out different kinds of questions. Select a subset of 10 of your questions to copy and paste into this Google document, along with the answers that you received. Include questions that you thought would be especially challenging, that START was able to answer to your satisfaction, as well as questions within the domains of knowledge that it claims to know, but which did not yield good answers. Complex questions may require multiple reasoning steps to sort out, for example, *Who is the president of the fourth largest country married to?* Add a brief comment about why you selected each example. Do you think START is intelligent? Why or why not?

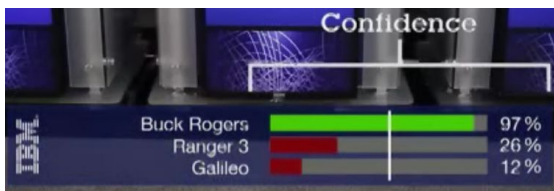


[IBM's Watson](#) was designed to be the world's most advanced, broadly knowledgeable question answering system, designed to meet the challenge of beating the best human players at the game of Jeopardy. You can read more about Watson in this 2010 [New York Times article](#).

Watch the following four segments of this [Watson video](#) showing snippets of Watson in action, playing Jeopardy against human contestants. As you are viewing the video, consider **Q2** below.

- (1) 11:55 - 17:55 (6 mins)
- (2) 22:57 - 25:23 (2.5 mins)
- (3) 26:04 - 28:00 (2 mins)
- (4) 28:12 - 34:00 (4 mins)

After each Jeopardy clue is revealed, a graphic appears on the screen that shows Watson's top three "answers" and an associated *confidence* in each of these answers, as shown below. The vertical white bar shows Watson's *buzz threshold*, which is the level of confidence needed for Watson to "hit the buzzer" to give an answer. This buzz threshold changes depending on the current game situation.



Q2. List at least 5 questions and answers here that Watson got right, which surprised or impressed you the most. You may think, for example, that these questions are very challenging because the phrasing of the Jeopardy clue is somewhat convoluted, or the knowledge needed to answer the questions is very esoteric, or that complex inferences are needed to construct the right answer. Also list at least 3 questions that Watson got wrong that you think are fairly straightforward questions, along with the answer that Watson gave. Add a brief comment about why you selected each question, including those that Watson got right and those that Watson answered incorrectly. For questions that Watson got wrong, why do you think people did better on them?



[home](#) • [chat to me](#) • [news](#) • [chatlogs](#) • [tips](#) • [contact us](#) • [awards](#) • [free aiml files](#) • [links](#) • [gallery](#)

Have a chat with the [Mitsuku Chatbot](#), winner of the [Loebner Prize](#) in 2013 and 2016.

Q3. Select several of your exchanges with Mitsuku and briefly say whether or not Mitsuku's replies seem like natural human responses, and why. Does Mitsuku seem intelligent to you? Why or why not?

Q4. What is the Loebner Prize? Be sure to include a brief description of the idea of a "Turing test" in your answer. Do you think a computer program will ever win the Gold Medal? Why or why not?



Use [Google Translate](#) with English and a second language that one or both of you knows well.

Q5. Enter some easy sentences and some difficult ones into Google Translate, and report on the outcome. Be sure to do translations in both directions, for example, from English to Spanish and also from Spanish back to English. Try to find one or more examples that Google Translate does not translate well, because it seems to lack certain commonsense knowledge to interpret the true meaning of the sentence.