

CSC247/447 Natural Language Understanding

Programming Assignment #1: Due Feb 26th

Your task is to build an interactive system that can accept text input and identify the correct intention underlying it, drawing from the four tasks shown in the sidebar. You should use the frame-based methods as described in the lecture, defining a frame (or a set of frames if you wish) for each task. Your completed system will be tested on a range of utterances, including some unusual ways of phrasing things, so you should define your patterns in terms of ontology types to bring in as many approximate synonyms as possible. To facilitate this, attached are files that provide the TRIPS lexicon (mapping words to their possible ontology types) and the TRIPS ontology (defining a hierarchy of ontology types). *To allow this, your system needs to read these files and implement a search over the ontology to enable the ontological matching.*

You also might consider using constraints to improve your selectivity (e.g., something matching a NUMBER could only be a time if it is between 1 and 12). The easiest way to browse the lexicon and ontology is using the web interface¹.

Your system will be tested by providing it with a file that contains a list of sentences, each one to be interpreted independently as the first utterance of an interaction. Your program should return a number of pieces of information for each sentence, as described here:

1. Whether the sentence is ambiguous wrt the task to be performed (SUCCEED/AMBIGUOUS/FAIL)

If the task is successfully identified

2. Which task is identified
3. Is the task ready to execute, or would the system need to come back with a clarification request (READY/CLARIFY)

If the task is ambiguous

2. A list of possible tasks that might be evoked

The following table gives examples of inputs and out-

THE TASKS AND EXAMPLES

Using Contacts (CONTACTS)

Looking up or adding contact info (phone #, and email address)

What is Mary's phone number
Give me Matt's email
Bob's number is 555 3333

Initiating a Phone Call (PHONE)

Using entry in the address book or an explicit number

Phone Matt
Call 576 6767

Email (EMAIL)

Reading email, or starting an email - do not have to handle adding the content of the email!

Start an email to John
Is there any mail from Matt
Send Lisa a note

Text Messaging (TEXT)

Reading a text msg, or starting a text msg- do not have to handle acquiring the content!

Text John
Any messages from John?

Simplifications for Evaluation

1. People referred to by first name only
2. We will only use the names specified in the examples above
3. Messages will only use the base form of words (no tense, number variants) E.g., instead of "Bob's" you'll get "Bob"; Instead of "messages" you'll get "message"
4. Phone numbers have no area codes

¹ www.cs.rochester.edu/research/trips/lexicon/browse-ont-lex-ajax.html

puts. If this was the test, you would get a file with these four sentences, one on each line, and you would output a file with the four answers, one on each line.

<i>Input</i>	<i>Response</i>
I want Matt email	SUCCEED CONTACTS READY
Contact Matt	AMBIGUOUS PHONE EMAIL TEXT
Text	SUCCEED TEXT CLARIFY
I am happy	FAIL

Table 1: Example inputs and outputs

Information Provided

The attached TRIPS ontology and lexicon files are in YAML format. There are libraries available in Java and Python for YAML as well as other languages. In addition, a third file gives mappings from WordNet synsets to TRIPS ontology types, which might be useful in a team effort where you try to integrate WordNet into the system to expand the vocabulary your system can handle.

Feel free to extend the ontology with additional types if you find them useful for your system, and describe such changes in your documentation.

Teams

The assignment is designed to be manageable for doing as an individual effort. If you wish, however, you can plan a more ambitious system as part of a team. Examples of things you might do to extend the core system are

- Incorporate WordNet to expand the vocabulary
- Close the loop and have the system generate responses and engage in a series of interactions to task completion
- Perform morphological analysis so that we can relax the constraint of only using base forms of the words
- Any other generalizations you can think of that would improve the system.

You are, of course, welcome to try these generalizations as an individual. But if you want to work in a team, you should send me your system development plan by **Feb 17th**. Your plan should discuss the generalizations you will be making and also include a system diagram that shows which part of the system each team member is responsible for.

Documentation

Documentation is as important as the code. Make sure you describe your representation of Frames to a level of detail so that someone else could be able to add a frame to your system. Also, provide us with description of the Frames you defined for you system, using an abstract notation such as that used in the lecture notes, or some other easy to understand format.

Questions

If you have questions on getting the project started, using the data, or the evaluation, please feel free to contact Rik Bose (rbose@cs.rochester.edu). Rik could also meet with you on Monday afternoon. Contact him if you'd like to meet.