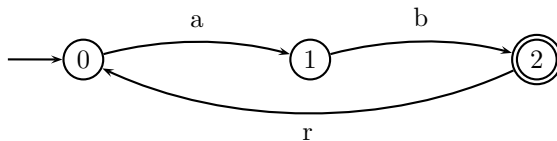


Assignment 7

Linguistics 445/515
Dialogue Systems & ICALL

Due Monday, December 8, 2008

1. There is a corpus of academic transcripts at <http://micase.umdl.umich.edu/> – the Michigan Corpus of Academic Spoken English (MiCASE). Pick one of the linguistics transcripts and answer the following questions:
 - (a) What do you notice about the turn-taking (length of time, interruptions, etc.), and what does this tell us about the relationship between the speakers?
 - (b) How do you see the speakers establishing common ground? Be specific.
 - (c) Do the following for the first four exchanges (i.e., about 8 turns):
 - Divide the speech into *utterances*
 - Label each utterance with at least one DAMSL tag
2. Say you have to design your own ELIZA-type system. The task for your ELIZA (or whatever you want to call him/her/it) is to be the host of a radio call-in show (on a topic of your choice). Sketch how you would design such a system, considering:
 - What kinds of input do you expect? Are there certain templates to include?
 - What words/phrase (stock phrases) would you be sure to include?
 - What words/phrases do you want to avoid?
 - What do you see as the biggest problem for your system? Lack of knowledge? Lack of memory? Lack of grammar? etc.
 - Would you expect your system to outperform a human?
3. (a) Draw a finite-state automaton (FSA) for the regular expression $sha(na)^*$
(b) Here's an FSA:



- i. What is the shortest string this FSA will generate?
 - ii. What are two other strings it will generate?
 - iii. Write a regular expression to capture this FSA.
4. You're going to be looking at the Monroe corpus, which is very similar to the TRIPS corpus—namely, two people are engaged in some sort of problem-solving task: <http://www.cs.rochester.edu/research/cisd/resources/monroe/s2/s2.transcript>
Using this corpus, find and explain:
 - (a) one pronoun (other than *I* or *you*) which is easy to resolve
 - (b) one pronoun which is difficult to resolve

You should include some description of how the pronouns are computationally resolved.
5. Interview someone taking a language course here at Indiana.
 - (a) How far along are they in their acquisition of this language?
 - (b) Identify a problematic area in their language learning.
 - (c) Design an exercise that would address their need. Consider the following in designing the exercise:
 - What kind of exercise best addresses their particular problem?
 - What kind of feedback do they prefer?
 - What level of users would this be appropriate for?
 - Can this be implemented on a computer, and if so, what technology is needed?
6. Try out the WERTi system online (<http://prospero.ling.ohio-state.edu/WERTi/>) and answer these questions:
 - (a) In a sentence or two, describe the positives of these types of exercises.
 - (b) In a sentence or two, describe the limitations of these types of exercises
 - (c) Describe in detail how a POS tagger can be used to generate fill-in-the-blank exercises. How could this be extended to other languages?

7. **Bonus:**

A lot of warnings about science and technology use are couched in science fiction terms. So, for example, *The Matrix* and *2001: A Space Odyssey* warn against dependence on computers; *Dr. Strangelove* warns against the production of atomic bombs; *The Terminator* highlights what happens when robots go bad. And so on. Unfortunately, rarely does an author rail against the dangers of language technology. We don't usually see a movie displaying the dangers of spell checkers or machine translation systems or whatever.

Your task is to sketch out a plot for a story warning against the use of some language technology that we have discussed this quarter. (The danger could range from bad student grades to larger unemployment to world destruction. Whatever.) Approximately half a page is sufficient, although longer stories are welcomed, if you feel so inclined. (And if you don't want to do any "creative" writing, then simply write an essay-style half-page on the dangers of some technology.)