

Homework 1: Eliza (133 points)

Please post any questions concerning this assignment to the Courseworks (<http://courseworks.columbia.edu>) discussion board, under the Homework 1 topic.

1 General Instructions

Eliza, (<http://www.manifestation.com/neurotoys/eliza.php3>) as demonstrated in class, imitates a Rogerian psychotherapist via simple pattern matching and substitution using regular expressions. The main goal of Homework 1 is to produce an Eliza with a distinct 'personality'. You will be graded on concept as well as execution.

Your Eliza must respond both to queries from a file, one line at a time, and take interactive console input. When responding to queries from a file, Eliza should read the file, respond to each line, and terminate; when responding to console input, it should loop until the user quits. We will provide a sample file as input for 'file mode'. You can use this file as 'development data' as you build your Eliza.

You must write the Eliza code yourself — don't use publicly available code (please refer to the Academic Integrity policy if you have questions in this regard). You may use NLP analysis tools available on the web (e.g. part-of-speech taggers, morphological analyzers, parsers) but you must indicate which ones you use, for what purpose, and where they were obtained. Unless you have discussed it with Frank beforehand, your assignment should be written in perl, Java, Python, c or C++ **and compile and run on the CS cluster machines**. Do yourself a favor and use a language that has nice text processing features.

Your submission must include a README file as specified in Section 2.2 below. Also include code for your program and any supporting data files you use, as well as the required compilation and execution scripts as described in Section 3 below. Your submission should not include compiled code.

2 Grading

You will be graded on the following elements:

2.1 Functionality (110 points total)

Novelty (40 pts.) How novel/interesting the personality you choose is and how effectively you implement it.

Correct grammar (30 pts.) How grammatically correct your system output is.
e.g. I AM SAD → WHY ARE YOU SAD? (not *WHY YOU SAD?).

History mechanism (20 pts.) You should exhibit some notion of memory of previous dialogue content on Eliza's part. For example, if the user begins by talking about his/her dog, Eliza may reintroduce that topic later on in the same dialogue.

Variability (20 pts.) Eliza should not respond the same way to the same input every time.

Overachieving (?? pts.) Any improvements above and beyond these will be considered for extra points, e.g. what novel features you use to analyze input or produce output.

2.2 Software Engineering (includes documentation) (23 pts.)

- Your README file must include the following:
 - Your name and email address.
 - A description of the personality you have designed for Eliza, a description of every file in your solution, the programming language used, supporting files, any NLP tools used, etc.
 - How your Eliza operates, in detail.
 - A description of special features (or limitations) of your Eliza.
- Within Code Documentation:
 - Methods/functions/procedures should be documented in a meaningful way. This can mean expressive function/variable names as well as explicit documentation.
- Coding Practices:
 - Informative method/procedure/function/variable names.
 - Efficient implementation
 - Programmer, Memory, and Processor efficiency — don't sacrifice one unless another is improved
 - Don't hardcode variable values, etc.

3 Submission instructions

If you use a language that requires compilation, you must include a shell script that automatically compiles your code **on the C.S. cluster machines**. This file should be called `make.sh`.

Regardless of language used, you must include a simple perl wrapper called

```
hw1.pl
```

so that your program can be run as follows (again, on the cluster machines):

```
$ perl hw1.pl <filename>
```

Note that `<filename>` is optional, as your Eliza must respond both to queries from a file, one line at a time, and take interactive console input.

Hint: this simple perl script lists the contents of the present working directory:

```
print `ls`
```

When you have completed your Eliza, you will submit your solution electronically using instructions that will be provided in the coming week.

4 Academic Integrity

Copying or paraphrasing someone's work (code included), or permitting your own work to be copied or paraphrased, even if only in part, is not allowed, and will result in an automatic grade of 0 for the entire assignment or exam in which the copying or paraphrasing was done. Your grade should reflect your own work. If you believe you are going to have trouble completing an assignment, please talk to the instructor or TA in advance of the due date.