**­CIS 263 Project 1**

**Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Due Date**

* at the start of class on Monday, March 9.

**Before Starting the Project**

* Read the entire project description before starting. This is an individual assignment.

**Learning Objectives**

After completing this project you should be able to implement basic search using trees.

**Rubric**

25 pts commented and readable code \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

25 pts elegant source code \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

25 pts concise design \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

25 pts results during testing \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Background**

William Shakespeare carefully reuses character names, places, and terms between his writings. He asks you to build a search engine to allow him to search his previous writings to help in writing his next play. You decide to store each of his plays in a separate tree and search the multiple trees as needed to answer the questions he has.

**Step 1: Gather data and store the trees**

* Download the three plays from blackboard (from Project Gutenberg). Assume these are the only files your program will be tested with.
* Read each play into your program and store each word (token/term) in the play into the tree for that play (three separate trees). You may use any type of tree desired, including trees from existing libraries. You can choose how to handle duplicates (recall the zipf law regarding the use of words in the English language). If you use an existing tree implementation someone else wrote, you should be able to answer questions on how the tree is structured and how its functions are implemented.

**Step 2: Functions**

* Write functions as needed to determine if a given term has been used in a specific play, i.e., which plays the term occurs in.
* Write functions as needed to determine how many of the three plays contained the given term.

**Step 3: Testing**

* Provide output for the following questions, after the three plays have been stored. Utilize your tree structure only, not the raw text. The questions can be hardcoded into your program.
  + Which plays, if any, contain the term ‘Hamlet’?
  + Which plays, if any, have a ‘friar’ in them?
  + Which plays, if any, have the term ‘Lady’ occur?
  + How many of the plays have the term ‘servant’?
  + How many of the plays discuss ‘Italy’?
  + How many of the plays discuss ‘England?

**Step 4: Bundle your program**

* Turn in all files required to run your program. Also, include the testing output (answers) from Step 3 of your program. Print and staple all of these files and turn them in at the start of class on the due date (along with the first page of this description). Upload to blackboard only the source code.
* Ensure you include suitable documentation for your 1) overall project and 2) source code. Include an appropriate comment block at the top of your program.

**Grading Criteria**

* There is a 50% penalty on programming projects if your solution does not execute or generates errors.
* There is a 50% penalty for not turning in a hardcopy (code and 1st page of this document) and softcopy (zipped if needed) to blackboard.
* Any options/approaches/requirements not specified in this document are left for your own decision making, in keeping with the spirit of the assignment.

**Late Policy**

Projects are due at the START of the class period and not accepted later. Not turning in the hard copy or soft copy by the due date is considered a late/missing project unless PRIOR arrangements are made.

**Example Output (exact format not required)**

[eos11:~/CIS263]$ /usr/local/gcc-4.8.1/bin/g++ project1.cpp -o project1 -std=c++11 -static

[eos11:~/CIS263]$ ./project1

Analyzing Hamlet, Romeo and Juliet, and The Tempest

...

Q3: Which plays, if any, contain the term Lady occur? Hamlet RomeoAndJuliet TheTempest

...