CPSC-315 – Programming Studio
Team Project 3: Information Retrieval and Visualization

Project Goal:
Your team is to build a system for tokenizing, stemming, indexing, and ranking ascii text files for the retrieval and visualization via a query/browsing interface. Your system will be given the path of a directory containing the text files in the collection. Your system will then provide a user interface for specifying a query and viewing the retrieved documents as a ranked list and a visual user interface for browsing the collection as a whole or a subset of the collection. The ranked list interface is expected to look much like most Internet search engines (e.g. Google). The visual user interface is where your team is expected to be creative. The visual user interface should be able to work for the whole collection and for a subset of the collection (e.g. the top N documents returned by a query).

Specifically, your team will create a number of components to achieve this goal:

- Tokenizer – takes a string and returns the words as tokens
- Stemmer – identifies the stem of a given word
- Inverted index – provides access to all of the instances of a stem in the document collection
- Ranking algorithm – rank orders the documents in a collection for a given query
- Query/ranking interface – simple query input and results interface
- Collection/results visualization – creative presentation of a set of documents based on similarity or ?

This is the basic functionality required. All of these aspects can be improved in numerous ways.

Note that the four-person team will have higher expectations in this application program than a three-person team would.

Team Organization:

Each project team has two or three members, one of which is the "project leader" and has additional responsibilities. Each month-long team project will consist of three phases:

Week 1: Project leader develops high-level specification of Java objects to be instantiated and identifies which team member will be responsible for each object. This specification takes the form of the names of the objects, their methods (and parameters), and any externally available constants or variables and comments describing each of the above.

Weeks 2-3: Team members implement their assigned software components, communicating as needed to ensure necessary changes to the original specification are known to all team members.

Week 4: All team members work to pull together their components into the final, working system and generate all necessary documentation for the project.

After the project: Each team member will be asked to assess the work of the other team members.

For this assignment, the teams will be as follows (project leaders are listed in italics):

Team 1:
Andrew Reagan
Jorge Cereijo-Perez
Christopher Bennett

Team 2:
Brett Hlavinka
Benjamin Unsworth
Drew Havard
Patrick Robinson

Team 3:
Zachary Edens
Robert Kern
Andrew Johnson

Team 4:
Ryan Mcauley
Jeffrey Deuel
Jillian Graczek

Team 5:
Mark Hill
Kourtney Kebodeaux
Brandon Jarratt

Team 6:
Gabriel Copley
Christopher Aikens
Julio Montero

Team 7:
Travis Kosarek
Thomas Robbins
Christopher Weldon

Team 8:
John Laky
Matthew Moss
Jacob Lillard

Code Organization:
This project is to be completed in Java.

The project leader is required to maintain its development using a version control system. You should choose and set up an SVN repository/project site in which your team will keep current versions of source code, documentation, etc. This should be one of the first decisions your team makes. CodePlex, Google Code, and Sourceforge are suggested locations for these repositories, but you may choose another option if you would prefer. Note that your team is to make use of this repository for all development. You should not pass around files by email, saving in shared directories, etc. Significant points will be taken off of the project if code is shared via a method other than the SVN system. (You may use a version control system other than SVN, if you would prefer).

Your project leader should give access to both the Instructor and the TA to the SVN repository for download. Your project leader is required to send email to both giving the location of your team's project, and instructions for access to the SVN repository (this should be done by the time of Project Update 1, below). The instructor/TA may download SVN software at any point.

Intermediate Updates:

Organization Statement: Your project leader is to turn in a written update regarding the overall project design. Specifically, your report should have 3 main components:

- A brief summary of what system you are using for version control, and a short justification (1/2 page) for why your team chose that option. By this point, you should have emailed the professor and TA with details on your SVN repository, and any details they need regarding access.
- A statement describing how your team will be organized and any responsibilities assigned to team members. You may use any organization, but you should give a justification/reasoning behind whatever you have chosen. This summary should be ½ to 1 page in length.
- A brief (1 to at most 2 pages) summary of the approach your team will be taking for implementation of the project. It is recommended that you try to identify the major implementation issues, set intermediate deadlines for the team (building in some slack time), including for production of documentation, and assign responsibilities to members (at minimum, for the initial work)

Project Grading:

The overall project will be graded as described below. This grade will be used to determine the team grade, and individual grades will be determined by apportioning the team grade among the team members.

15% Final Application Documentation
40% Tokenizing/Stemming/Indexing/Ranking of documents
35% Browsing/Visualization of collection and search results
10% Code style: naming/layout/commenting

The project leader will be graded as described below.

20% Source system selection & explanation of use
30% Design of Application – How complete is the design? Is it specified clearly for team members.
30% Description of Team Organization (who is responsible for what)
20% How well the team worked (from individual reports)

Dates and Deadlines:

The following are the intermediate deadlines for this project. Details of the particular requirements are described above. Note that these are final deadlines; it is advisable that your team leader complete the leader report in advance of the "due date".

Tuesday, April 2
Project Assigned
Tuesday, April 9, 11:59 p.m.
Email Instructor and TA regarding SVN repository
Project Leader Report (source system, team organization, and API/application design/code stumps)
Thursday, April 30, 11:59 p.m.
All final project code turned in
Project documentation turned in along with code

One day after code and documentation turned in (or **Friday, May 1**, 11:59 p.m.)
Individual Reports on teamwork turned in