

Course Syllabus for STAT 527: *Introduction to Text Mining* Spring Semester, 2012



Professor: Roger Bilisoly,
Associate Professor of Statistics
Central Connecticut State University

Office: Marcus White 112
Phone: (860) 832-2852
Email: bilisolyr@ccsu.edu or class email

Required Texts:

Practical Text Mining with Perl

by Roger Bilisoly

ISBN: 978-0-470-17643-6

SAMS Teach Yourself Perl in 21 Days

by Laura Lemay

ISBN: 0-672-32035-5

Please purchase these two books as soon as possible. You will need both to do the online quizzes, and there will not be extensions due to not having these books.

Required Software:

Perl (for Linux/Unix) or ActivePerl (for Windows) or MacPerl (for the Macintosh). Perl is free and is available under the Artistic license. To download any of these versions, go to <http://www.perl.com/download.csp> and follow the directions given there. Note that I am only familiar with using Perl on Windows.

The use of Clementine for this class is optional. This is freely available to CCSU students via our Citrix server. For help on using Citrix contact the CCSU IT Helpdesk at 860-832-1720. For the Edgar Allan Poe final project, Clementine is not needed, but it could be used for a proposed final project.

Prerequisites and Requirements:

The prerequisite for this course is STAT 521 or permission of the instructor. Since this class will be entirely online, you need to be mature and able to work independently. You'll have to become familiar and comfortable with Blackboard Vista, the software that makes the class available over the Web. This course requires extensive use of the programming language Perl. Any programming language takes effort to learn, **so finishing work on time is essential**. To encourage turning in materials on time, **late submissions won't be accepted, and they will get a grade of 0%**. Please don't wait until just before the deadline to submit class materials: having a computer problem just before a due date is **not** an acceptable excuse for a late submission.

Please note that **the final project must be finished by its due date, otherwise you will fail the course.** Moreover, there will be no incompletes for this course. If procrastination is a problem for you, then don't take STAT 527.

How Will This Course Work?

There will be seven lessons and six online quizzes. You will get two chances to take each quiz and only the better score will be used. All together, the quizzes make up 50% of your grade, and the final project makes up the other 50%. Below are the due dates:

- Do Lesson 1 as early as possible: at the very latest by Wednesday, January 25. Note that there is no online quiz for Lesson 1.
- Lesson 2 and Online Quiz 2 due Wednesday, February 1
- Lesson 3 and Online Quiz 3 due Wednesday, February 8
- Lesson 4 and Online Quiz 4 due Wednesday, February 15
- Lesson 5 and Online Quiz 5 due Wednesday, February 29
- Lesson 6 and Online Quiz 6 due Wednesday, March 14
- Lesson 7 and Online Quiz 7 due Wednesday, April 4
- If you choose to propose a final project, submit the proposal by Wednesday, March 28
- I will make a decision on the final project proposal by Wednesday, April 4
- Final project is due Wednesday, May 2. Remember that there are no extensions on the final project.
- There will be a data collection assignment, due date to be announced.

All submissions are due 11 pm CCSU time (US Eastern Time), which may **not** be your local time.

Final Project:

The Final Project is worth half of your grade, so doing it well is essential for a good grade in this course. The objective of this project is to compare the style of Edgar Allan Poe's short stories both among themselves and with short stories by other authors. A detailed project description will be posted during the semester.

Getting Help:

Your fellow students are a valuable source of help and advice, so please use the discussion threads for this course, both to ask and to answer questions. If you wish to talk to me over the phone, we can set up a time to talk via email. In general, I will check my email each weekday, but neither over the weekend, nor on CCSU holidays, nor during spring break.

If you have problems programming Perl, please post to the appropriate class thread and include the portion of code that is giving you problems.

If you have computer problems, please contact the CCSU Helpdesk at 860-832-1720.

University's Academic Integrity Policy:

All students are expected to demonstrate academic integrity in the completion of their coursework. This includes doing one's own work and giving proper credit to the work and ideas of others. Students who engage in plagiarism and other forms of academic misconduct will face academic and possibly disciplinary consequences. Academic sanctions can range from a reduced grade for the assignment to a failing grade for the course. From a disciplinary standpoint, an Academic Misconduct Report may be filed and a Faculty Hearing Board may impose sanctions such as probation, suspension or expulsion.

It is your responsibility to know the Student Code of Conduct (see <http://www.ccsu.edu/page.cfm?p=4155>) and the Policy on Academic Misconduct (see <http://www.ccsu.edu/page.cfm?p=2468#Misconduct>.)

Drop Date:

The last day to drop a course is **April 17**. Approvals for withdrawal are not required if done by this date; however, it is strongly recommended that students consult with their academic advisors prior to deciding to withdraw. Cessation of participation, notice to the instructor, or telephone calls to the Enrollment Center are not considered official notice of a student's intention to drop the course.

After the above date, withdrawals are allowed only **under extenuating circumstances** and require approval of the course instructor and department chair.