

# CSCI 493

## Mentoring: Stylometry

Fall 2016

Tu 11:15am -1:05pm

SCI 168

<http://cs.wheaton.edu/~tvandrun/cs493>

**Thomas VanDrunen**

☎630-752-5692

☎630-639-2255

✉Thomas.VanDrunen@wheaton.edu

Office: SCI 163

Office hours: MTuWF 3:30–4:30pm;  
Th 9:30–10:30am and 1:00–3:00pm.

---

### *Contents*

**CATALOG DESCRIPTION.** Faculty and student collaboration on a project of mutual interest.

**ADVERTISED DESCRIPTION** Stylometry is the statistical study of an author's style, measuring various stylistic attributes to form statistical models of an author's style or voice. It can be used to determine authorship in an anonymous or disputed work, to observe changes in an author's style over time, to quantify how one author has been influenced by another, or to determine the relative contributions of different authors in collaborative works.

We will read literature on recent advances in the field of stylometry and explore the use of available software (including the Stylo package for the R system). This will require us to learn some probability, statistics, and machine learning, including learning the R programming language for statistical analysis. We will then experiment on real world texts to discover evidence for authorship or influence.

**OBJECTIVES.** The objectives of this course are to learn the methods of stylometry, the available tools, the statistical background, and the research that stylometry has been applied to.

At the end of the course, students will be able to

- use Python to process texts, including use of the `nltk` library;
- use R to process statistical data about stylistic markers, including use of the `stylo` package;
- articulate the statistical principles used in computing information about an author's style;
- analyze a text for evidence about authorship, collaboration, influence, and other questions related to authorship style.

### *Course procedures*

**HOW WE DO THIS COURSE.** This course is low-structure. At each meeting we will have devotions, discuss a paper, and report on our projects (once projects have started). Many topics in the course will be thing I myself am studying for the first time, and I plan to do a project like yours myself<sup>1</sup>. It will be a collaborative learning experience.

**READINGS.** I will make the readings available through Schoology. Please *print them out*, read them, and bring them to class, ready to discuss them.

**TERM PROJECT.** You will do one big term project which will involve studying a text or texts using the methods and tools we learn about in the papers. Included in your term project will be a publication-quality paper describing your problem, methodology, and results.

---

<sup>1</sup>instructor reserves the right not to do a project

**FINAL MEETING.** The final exam block for this period is Thursday, Dec 15, 1:30-3:30 pm. There will not be a final exam in this course, but we are required to hold a final meeting.

**GRADING.** Your grade in this class will be based on

<i>instrument</i>	<i>weight</i>
Participation	25
Project	75

“Participation” entails doing the readings, contributing to class discussion, and completing minor exercises and assignments as they come up.

### *Policies etc*

**ACADEMIC INTEGRITY.** Any occurrence of plagiarism will be handled according to the college’s policy.

**ATTENDANCE.** Students are expected to attend all class periods. It is courtesy to inform the instructor when a class must be missed.

**EXAMINATIONS.** There are no examinations in this course.

**SPECIAL NEEDS.** *Institutional statement:* Wheaton College is committed to providing reasonable accommodations for students with disabilities. Any student with a documented disability needing academic adjustments is requested to contact the Academic and Disability Services Office as early in the semester as possible. Please call 630.752.5941 or send an e-mail to [jennifer.nicodem@wheaton.edu](mailto:jennifer.nicodem@wheaton.edu) for further information.

*My own statement:* Whenever possible, classroom activities and testing procedures will be adjusted to respond to requests for accommodation by students with disabilities who have documented their situation with the registrar and who have arranged to have the documentation forwarded to the course instructor. Computer Science students who need special adjustments made to computer hardware or software in order to facilitate their participation must also document their needs with the registrar in advance before any accommodation will be attempted.

**GENDER-INCLUSIVE LANGUAGE.** The college requires the following statement to be included on all syllabi: *For academic discourse, spoken and written, the faculty expects students to use gender inclusive language for human beings.*

**OFFICE HOURS.** I try to keep a balance: Stop by anytime, but prefer my scheduled office hours. Any time my door is closed, it means I’m doing something uninteruptible, such as making an important phone call. Do not bother knocking; instead, come back in a few minutes or send me an email.

**DRESS AND DEPARTMENT.** Please dress in a way that shows you take class seriously—more like a job than a slumber party. I feel this is especially important to mention since this is a morning class. Moreover, since this is a discussion class, you should note that how you present yourself communicates to your classmates how seriously you take the class, and how seriously you take them. (If you need to wear athletic clothes because of activities before or after class, that’s ok, but try to make yourself as professional-looking as possible.)

**BRINGING READINGS TO CLASS.** Bring a **hard copy** of the day’s readings to class with you. Horay for paper.

**ELECTRONIC DEVICES.** Please keep laptops and all electronic devices put away and silenced during class. (That’s right, this is a computer science course, but you’re not allowed to use a computer during class. Trying out programming concepts on your own during class time (for example) is not productive because it takes you away from class discussion. You cannot multi-task as well as you think you can.) ***Text in class and DIE.*** I take this very seriously.