

**DISC 101 – Introduction to Discovery Informatics**  
**Fall 2008**  
**Syllabus**

**Instructor:** Semmy Purewal, Ph.D. ([www.cs.cofc.edu/~semmy](http://www.cs.cofc.edu/~semmy))  
**Contact:** [semmy@cs.cofc.edu](mailto:semmy@cs.cofc.edu), (843)-953-7542  
**Office:** JC Long 219  
**Office Hours:** MWF 12:00-3:00, also available by appointment

**Class Time:** MWF 9:00-9:50, JC Long 221  
**Course Homepage:** WebCT  
**Text:** *Ambient Findability* by Peter Morville.

**Course Description (College Catalog):** Introduction to knowledge discovery techniques, emphasizing computer based tools for the analysis of large data sets. Topics include the discovery informatics process and inductive data-driven modeling. Students will have hands-on experience with Online Analytical Processing (OLAP) and data mining software and complete a project.

**Course Description (My Description):** Broadly speaking, we'll be learning how to use computers to cope in an age of 'information overload.' More specifically, we'll survey the fundamental ideas behind the automated knowledge discovery process. This will include information representation and storage, computation and algorithmic thinking, and automated learning techniques.

**Evaluation:** You'll be evaluated based on your performance on tests, homework/projects, and various other activities described below.

Exams: We'll have three in-class, closed book/closed notes exams that will each account for 15% of your grade. I'll describe the format in class before the exam date. In order to pass the class, you must take the final exam.

Homework/Projects: Throughout the semester, we will have various homework problem sets or project-type activities. The value of a single homework assignment or project will depend on the total number of activities given throughout the semester. The average will account for 20% of your grade.

In-class Activities: We'll have various unannounced in-class activities throughout the semester. Some of these will have a component that is turned in, and some will not. When you turn something in with your name on it, it will count toward your 'in-class activities' grade. The value of a single in-class activity will depend on the total number of activities given throughout the semester. The average will account for 20% of your grade.

Technology Breadth Requirement: Since this is an introductory course that is oriented around technology, we'll be required to participate in technology-oriented activities on campus or around Charleston. This can include, for example, attending meetings of a local tech group, or starting/participating in a community service project. Please drop by during office hours and we can find a way to meet this requirement that fits your interests. This will account for 10% of your grade.

Participation: This portion of our grade will include attendance (which will be taken randomly), discussion participation and participating in activities. This will account for 5% of your grade.

Grade Computation: I follow a relatively standard grading scale, meaning 90 guarantees at least an A, 80 guarantees at least a B, 70 guarantees at least a C, 65 guarantees a D, and below 65 is an F. Plus and minus grades may be given out at my discretion. Your numerical grade will be computed as follows:

Exam 1: 15%  
Exam 2: 15%  
Homework/Projects: 20%  
In-class activities: 20%  
Technology Breadth Requirement: 10%  
Participation: 5%  
Final Exam: 15%

### **Expectations and Policies:**

Attendance Expectations: I expect you to attend every class. If you are not going to be able to make it, I hope that you will drop me an e-mail letting me know. Attendance will be taken randomly and will count toward your participation grade.

Outside-of-class Expectations: I expect you to spend between approximately 5-6 hours per week outside of class on this material.

Class Expectations: I expect you to be attentive in class (no newspapers, crosswords, books, etc.) and that you participate in class discussions. I do not require that you raise your hand, but I do expect that you make an effort to interject at appropriate times, and that you are respectful and willing to yield the floor to other students.

Reading Expectations: I expect that the assigned readings be completed before the class in which the material is covered. Expect to spend at least 2-3 hours per week reading.

Exam Expectations: I expect you to be present for exams. As a general rule, I do not give make-up exams. If a serious issue arises, however, let me know ASAP and we can work something out. I can often give exams early if you have a college-approved reason that you have to miss them.

Technology Policy: If you would like to use a laptop or any other electronic device during class time, you must discuss it with me first. I prefer that you begin by writing a short note explaining how you will use it and how it will enhance your learning. Then we will come up with a plan whereby you will be held accountable for using it in the agreed-upon manner.

Homework Expectations: Homework expectations will vary. Sometimes you will be expected to work individually, and other times you will be expected to work with other students. By default, however, I expect you to work individually on assignments.

Academic Honesty Expectations: You are expected to abide by the honor code described on page 10 of the student handbook. When in doubt, please ask me whether than violating academic honesty.

Special Needs: If you have any special needs, please contact me as soon as possible via e-mail or office hours. I am happy and willing to accommodate all documented disabilities.

## **Tentative/Potential Topic List:**

### Introduction

0. Syllabus, Course Requirements
1. What is Discovery Informatics?
2. What are our goals for this class?

### Information

0. What is Information?
1. Information Theory
2. Binary Numbering System, ASCII encoding
3. Compression Schemes (Huffman Encoding)
4. Data Storage
5. Databases/SQL
6. Where does Data come from?
7. Technologies that generate information (Facebook, Last.fm)

### Computation

0. What is computation?
1. Algorithms/Algorithmic Thinking
2. Programming (Python, maybe Perl)
3. Computational Statistics (algorithms for basic statistical properties)
4. Foundations of Computation: DFA's, Turing Machines
5. Computability
6. Computational Complexity, Running Times
7. Sorting algorithms
8. P vs. NP
9. Artificial Intelligence

### Learning

0. What is learning?
1. Human Learning vs. Automated Learning
2. Statistical Learning
3. Machine Learning
4. Data Mining
5. Technologies that assist in Knowledge Discovery