

## COURSE DESCRIPTION

Dept., Number	CSci 345	Course Title	Information Storage and Retrieval
Semester hours	3	Course Coordinator	Yixin Chen, Assistant Professor

### Current Catalog Description

Examination of systems for storage and retrieval of information in textual and other formats. The topics include query processing, matching and ranking algorithms, text analysis, user interfaces, and evaluation of retrieval effectiveness.

### Textbook

C. D. Manning, P. Raghavan, and H. Schütze, *An Introduction to Information Retrieval*, Cambridge University Press, 2008.

### References

<http://informationretrieval.org>

### Course Outcomes

Upon successful completion of this course, the students:

1. can explain the basic structure and components of information retrieval (IR) systems,
2. are able to develop basic IR algorithms,
3. understand the Web search, Web crawling, and ranking algorithms.

### Relationship between Course Outcomes and Program Outcomes

This is an elective course for BSCS students.

The course outcomes contribute to the program outcomes as follows: (1) to (a), (2) to (b) and (c), and (3) to (a).

### Prerequisites by Topic

1. Basic data structure and algorithms (CSCI 211)

## Major Topics Covered in the Course

1. Information retrieval using the Boolean model. (2 hour)
2. The dictionary and posting lists. (2 hours)
3. Tolerant retrieval. (3 hours)
4. Index construction and compression. (3 hours)
5. Vector space models. (3 hours)
6. Similarity scores and ranking. (3 hours)
7. System evaluation. (2 hours)
8. Relevance feedback. (2 hours)
9. Probabilistic information retrieval. (2 hours)
10. Text classification. (3 hours)
11. Vector space classification. (2 hours)
12. Clustering (3 hours)
13. Web search (3 hours)
14. Link analysis (3 hours)
15. Exams. (3 hours)

## Assessment Plan for the Course

This is an elective course offered approximately every two years. An offering typically has 3 examinations and 9 programming and homework assignments. Outcome 1 is assessed by the exams and homework assignments, outcome 2 by the two programming assignments, and outcome 3 by the exams and homework assignments.

How Data in the Course are Used to Assess Program Outcomes (unless adequately covered already in the assessment discussion under Criterion 4)

--

## Estimate Curriculum Category Content (Semester hours)

Area	Core	Advanced	Area	Core	Advanced
Algorithms		2	Software design		
Data structures		1	Concepts of programming languages		