Master Course Syllabus

School of Engineering and Computer Science
Washington State University Vancouver

CS 451

Web Data Management

3 Semester Hours

Catalog Description

Introduction to concepts, data models, query and retrieval languages, and implementation issues for management of web data.

Prerequisite Courses

CS 351 – Introduction to database systems

Prerequisite Topics

- Relational data model
- Relational algebra and SQL
- Basic discrete mathematics
- Basic logic
- Use of Unix environment for coding, compilation, debugging and testing

Measured Course Outcomes

Students taking this course will:

1. Analyze or solve problems related to centralized and distributed indexing data structures and algorithms using suitable mathematics *(Contributes to performance criterion **A-3**)*
2. Evaluate the merits of multiple solution designs in web data management systems with respect to complexity *(Contributes to performance criterion **J-3**)*

Required Textbooks

Croft, Metzler, and Strohman. Search Engines: Information Retrieval in Practice, Addison Wesley, 2010. (ISBN 978-0-13-607224-9)

Reference Material

None Specified.

Major Topics Covered in the Course

1. Classic information retrieval concepts
2. Web information retrieval concepts
3. XML data model

4. XML query languages
5. Semantic web data models and query languages
6. Mining, crawling, and searching the web

Laboratory Projects

Programming Project Area	Weeks
Design and implementation of classic information retrieval systems	3
Design and implementation of web information retrieval systems	3
Design and implementation of web mining, crawling, or searching	3

CSAB Category Content

	FUNDAMENTAL	ADVANCED		FUNDAMENTAL	ADVANCED
Data Structures	0	1	Computer Organization and Architecture	0	.5
Algorithm & Software Design	0	1	Concepts of Programming Languages	0	.5

Oral and Written Communications

There are no significant oral or written communications required in this course.

Social and Ethical Issues

This course contains no significant coverage of social and ethical issues beyond the usual proscriptions against plagiarism and cheating.

Theoretical Content

Topic	Hours
Query algebras	6
Complexity analysis	2

Problem Analysis

Students take real-world web data management problems and implement efficient solutions. The problem analysis has two related components. The first component is determining the data management needs encountered in a particular contemporary web scenario. Students learn the capabilities and limitations of current techiques, through application design and implementation. The second component is an analysis of the practical limitations of current web data management solutions.

Solution Design

Students are guided through the several design phases in the development of a web data management system. Students analyze their solution to identify weaknesses in the design and implementation.

CC2001

This course provides coverage of topics in the following areas (hours listed are minimums):

IM1. Information models and systems [core]	9
IM2. Database systems [core]	9
IM3. Data modeling [core]	7
IM5. Database query languages [elective]	12

Course Coordinator: George Fletcher
Last Updated: April 22, 2009 (Approved)
Syllabus Version Number: 1.0