LONG SIGNATURE SHEET

UNC CHARLOTTE

Proposal Number: ITIS 11-23-10

Proposal Title: Web Mining (ITIS 5510)

Originating Department: Software and Information Systems

TYPE OF PROPOSAL: UNDERGRADUATE_____ **GRADUATE** __X__ **UNDERGRADUATE & GRADUATE**_____
(Separate proposals sent to UCCC and Grad. Council)

DATE RECEIVED	DATE CONSIDERED	DATE FORWARDED	ACTION	SIGNATURES
11/18/11	11/18/11	11/18/11	Approved	**DEPARTMENT CHAIR** Dr. Bill Chu
11/18/11	11/18/11	11/18/11	Approved	**COLLEGE CURRICULUM COMMITTEE CHAIR** Dr. Anthony Fodor
11/18/11	11/18/11	11/18/11	Approved	**COLLEGE FACULTY CHAIR (if applicable)** Dr. Barry Wilkinson
11/18/11	11/18/11	11/18/11	Approved	**COLLEGE DEAN** Dr. Teresa Dahlberg
			Approved	**GENERAL EDUCATION** (for General Education courses)
			Approved	**UNDERGRADUATE COURSE & CURRICULUM COMMITTEE CHAIR** (for undergraduate courses)
12-5-11	2-7-12	10-2-12	Approved	**GRADUATE COUNCIL CHAIR** (for graduate courses)
			Approved	**FACULTY GOVERNANCE ASSISTANT** (Faculty Council approval on Consent Calendar)
				FACULTY EXECUTIVE COMMITTEE (if decision is appealed)

Revised 08/01/11
OAA/lz

UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE

NEW GRADUATE AND UNDERGRADUATE   **ITIS 11-23-10**
COURSE PROPOSAL

FROM: DEPARTMENT OF SOFTWARE AND INFORMATION SYSTEMS

TITLE: WEB MINING 4510/5510

A. PROPOSAL SUMMARY AND CATALOG COPY:

1. SUMMARY:

The Department of Software and Information Systems proposes to create new courses, ITIS 4510 Web Mining, and ITIS 5510 Web Mining. The courses focus on both basic and advanced techniques of mining text-based information systems.

2. PROPOSED CATALOG COPY:

ITIS 4510. Web Mining. (3) **Prerequisites or Corequisites: ITCS 3160, or permission of department;** Topics include measuring and modeling the Web; crawling, Web search and information retrieval; unsupervised learning, supervised learning, semi-supervised learning in Web context; social network analysis and hyperlink analysis; text parsing and knowledge representation. (Spring)

ITIS 5510 Web Mining. (3) **Prerequisites or Corequisites: ITIS 5160, full graduate standing, or permission of department;** Topics include measuring and modeling the Web; crawling, Web search and information retrieval; unsupervised learning, supervised learning, semi-supervised learning in Web context; social network analysis and hyperlink analysis; text parsing and knowledge representation. (Spring)

B. JUSTIFICATION

1. Need

The Web has become a vast storehouse of rich data comprising text, hypertext markup, hyperlinks, sites, and topic directories. This course focuses on the application of data mining and statistical analysis techniques to discover patterns from the Web and mainly covers three sub-areas: Web usage mining, Web content mining and Web structure mining.

2. Prerequisites/Corequisites:

1. ITCS 3160/ITIS 5160, or permission of department.
2. Students should have some prior background in databases and statistics.

3. Course numbering:
ITIS 5510 is intended to be a graduate level course. ITIS 4510 is intended to be an undergraduate level course for senior students in Information Systems.

4. Effect on scope, quality, and efficiency:
The proposed courses will broaden the scope of the software and information systems curriculum to include the study of web mining. These crucial topics are destined to be even more important for our graduates to understand. Future IT professionals must understand how to apply a variety data mining techniques to web-based data for the purpose of learning or extracting knowledge.

C. IMPACT
1. Students served:
This courses provide graduate majors in information technology with an option for an elective course that would enhance their value to potential employers.

2. Effect on existing courses and curricula:
a. ITIS 4510/5510 will be offered in the spring.
b. The content/frequency of other courses will not be affected.
c. The anticipated enrollment is 20/20 students for ITIS 4510/5510.
d. Impact on enrollments in other elective courses will be minimal as it will be an elective course and the department offers a set number of electives each year
e. The proposed topics have not been covered in existing courses.
f. Other areas of catalog copy affected: The proposed courses should be listed as elective options.

D. RESOURCES REQUIRED TO SUPPORT PROPOSAL

1. Personnel
a. Specify requirements for new faculty, part-time teaching, student assistant and/or increased load on present faculty.

No new faculty members are needed in order to teach these courses.

b. List by name qualified faculty members interested in teaching the course(s).

In general, research faculty in data mining and data analytics areas are qualified to offer these courses. In particular, Xintao Wu, Mirsad Hadzikadic, Zibigniew Ras, and Wensheng Wu are interested in offering these courses.

2. Physical Facility

No new physical facilities are needed.

3. Equipment and Supplies

No new equipment and supplies are needed to teach the courses.

4. Computer
Specify requirements for computer usage by students and/or faculty, and include an assessment of the adequacy of computing resources by Computing Services.

Students will use the computing facilities in the CCI labs to complete their course projects and assignments. These facilities are adequate for the course.

5. Audio-Visual
Specify requirements for audio and/or visual equipment and media production services from Media Services.

This course requires only the use of existing presentation equipment in classrooms. No additional audio/visual equipment or services are needed.

6. Other Resources
Specify and estimate cost of other new/added resources required, e.g., travel, communication, printing and binding.

This course does not require any additional resources.

7. Funding Sources
Indicate source(s) of funding for new/additional resources required to support this proposal.

This course does not require any additional resources.

E. CONSULTATION WITH THE LIBRARY AND OTHER DEPARTMENTS OR UNITS

1. Library Consultation
Library consultation was initiated on 11-22-2010 and completed on 11-23-2010.

2. Consultation with other departments or units
Consultation with the departments of Computer Science and Bioinformatics and Genomics was initiated on 12-08-2010.

F. INITIATION AND CONSIDERATION OF THE PROPOSAL

1. Originating Unit
Approved by the Software and Information Systems Faculty on 11-23-2010 and by the College of Computing and Informatics faculty on 3-15-2011.

2. Other Considering Units

Consultation with the department of Computer Science as completed on 2/8/2011. Consultation with the department of Bioinformatics and Genomics was completed on 12-11-2010.

G. **ATTACHMENTS**

  1. ITIS 4510/5510 syllabus
  2. ITIS 4510/5510 schedule (includes list of topics covered)
  3. Library Consultation

Attachment A. ITIS 4510/5510 Course Outline

Course Catalog Description

ITIS 4510. Web Mining. (3) **Prerequisites or Corequisites: ITCS 3160 or permission of department;** Topics include measuring and modeling the Web; crawling, Web search and information retrieval; unsupervised learning, supervised learning, semi-supervised learning in Web context; social network analysis and hyperlink analysis; text parsing and knowledge representation. (Spring)

ITIS 5510. Web Mining. (3) **Prerequisites or Corequisites: ITIS 5160, full graduate standing, or permission of department;** Topics include measuring and modeling the Web; crawling, Web search and information retrieval; unsupervised learning, supervised learning, semi-supervised learning in Web context; social network analysis and hyperlink analysis; text parsing and knowledge representation. (Spring)

Prerequisites

ITCS 3160 or ITIS 5160, or permission of department.

Instructional Method

This course incorporates lectures and discussions of assigned readings. Graduate students in ITIS 5510 are required to study advanced topics such as opinion mining, sentiment analysis, or large graph mining; and to implement one state-of-the-art algorithm in the Web data analysis project.

Grading

All students will be evaluated by a course project, a midterm exam, and a final exam. Attendance and class participation may be factored into the final grade.

Term project and presentation	30%
Midterm exam	30%
Final exam	40%

Grading Policy:

for ITIS 5510, A = 90-100   B = 80-89   C = 60-79   U = below 60
for ITIS 4510, A = 90-100   B = 80-89   C = 70-79   D = 60-69   U = below 60

Academic Integrity

Unless otherwise specified, the UNC Charlotte guidelines on Academic Integrity specified under http://www.legal.uncc.edu/policies/ps-105.html fully apply to all work in this course. This includes term project and exams.

Textbook and resources

Web Data Mining: Exploring Hyperlinks, Contents and Usage Data, by Bing Liu, Springer, 2007. ISBN-13: 978-3-540-37881-5

Mining the Web: Discovering Knowledge from Hypertext Data, by Soumen Chakrabarti, Morgan Kaufmann, 2002, ISBN: 1-55860-754-4

Topics Outline

Week	Topics Covered
1	Introduction to Web mining and ethics issues in Web Mining
2	Association rules and sequential patterns
3	Supervised learning
4	Unsupervising learning and semi-supervising learning
5	Information retrieval and web search
6	Link analysis
7	Midterm
8	Web crawling
9	Information integration
10	Opinion mining
11	Web usage mining
12	Student project presentations
13	Social network analysis
14	Final

J. Murrey Atkins Library

Consultation on Library Holdings

To: Xintao Wu
Department of Software and Information Systems
College of Computing and Informatics

From: Reese Manceaux

Date: November 23, 2010

Subject: **ITCS 4510/5510 --- Web Mining**

Summary of Librarian's Evaluation of Holdings:

Evaluator: Reese A. Manceaux

Check One: Holdings are superior
              ☑   **Holdings are adequate (Please see comments)**   **YES**
                  Holdings are adequate only if Dept. purchases additional items.
                  Holdings are inadequate

Comments:

This is a proposal for an new course to learn both basic and advanced techniques of mining text-based information systems. It focuses on the application of data mining and statistical analysis techniques to discover patterns from the Web.

A small sampling of subject searching in the Atkins Library online catalog reveals the following holdings in support of these courses. (See the table that follows). A search of the related subjects retrieved over 3,000 pertinent items.

The Library has electronic access to periodicals and other electronic resources (e-books from NetLibrary & Skillport) that support these courses. Skillport, in particular, has an enormous catalog of computer related literature; especially up-to-date programming language books. In addition, the library has many electronic databases such as EBSCO databases (Cinahl), Springer Link, ACM Digital Library, IEEE Explore, ScienceDirect and Compendex (many with links to full

text articles) supporting the overall Computing and Informatics program. The collection, especially if supported by ongoing purchases, is quite adequate to support this program.

Atkins Library Holdings in Areas Related to

Web Mining

Library of Congress Subject Headings	Books	After Year 2000	Journals	Skillport /Books 24x7
Data Mining	163	106	6	680
Web Databases	13	12		187
Hypertext Systems	69	9	4	81
Social Networks – Mathematical models	8	4	1	99
Information Retrieval	152	37	9	913
Web Search Engines	39	1		438
Internet Searching (incl. Mathematics and Study)	37	24		193
TOTAL	**481**	**193**	**20**	**2591**

Reese A. Manceaux

Evaluator's Signature

November 23, 2010