UNIVERSITY OF MARY WASHINGTON -- NEW COURSE PROPOSAL

Electronically submit this completed form with PDF attachments to the Chair of the College Curriculum Committee.

COLLEGE (check one): [ ] Arts and Sciences [ ] Business [X] Education

Proposal Submitted By: Chris Garcia Date Prepared: 10/23/2012

Course Title: Analytics Application Development

Department/discipline and course number*: COB BUAD 400

*This course number must be approved by the Office of the Registrar before the proposal is submitted.

Number of credits proposed: 4 Prerequisites: Grade of C or better in CPSC 220 or equivalent

Will this be a new, repeatable “special topics” course? (Do you want students to be able to take this new course more than once if the topic changes?) [ ] NO [X] YES

Date of first offering of this new course: FALL SEMESTER, year [ ] Spring 2014

Proposed frequency of offering of the course: Once per Year

List the faculty who will likely teach the course: Chris Garcia

Are ANY new resources required? [ ] NO [X] YES Document in attached impact statement

This new course will be (check all that apply):

- Required in the major
- General Elective
- Elective in the minor [X] General Education**

**AFTER the new course is approved, a separate proposal must be sent to the General Education Committee.

Catalog Description:
ANLY/BUAD 400 - Analytics Application Development (4) - Prerequisite: CPSC 220 or equivalent. A course in programming and data manipulation techniques for constructing analytics-based applications. Topics include SQL and no-SQL databases, using web service API’s to acquire data, introduction to Hadoop and MapReduce, and use of third-party analytic component API’s. Cross-listed as BUAD 400.

COURSE HISTORY

Was this course taught previously as a topics or experimental course? [X] YES [ ] NO

Course Number and Title of Previous Course

<table>
<thead>
<tr>
<th>Course Number and Title of Previous Course</th>
<th>Semester Offered</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CHECK HERE if the proposed course is to be equated with the earlier topics or experimental offerings. This means that students who took the earlier “topics” course will only be able to take the new course if they made a C- grade or lower in the earlier course.

NOTE: If the proposed course has not been previously offered as a topics or experimental course, explain in the attached rationale statement why the course should be adopted even though it has not been tried out.

REQUIRED ATTACHMENTS:
1. Rationale Statement (Why is this course needed? What purposes will it serve?)
2. Impact Statement (Provide details about the Library, space, budget, and technology impacts created by adding this new course. Include supporting statements from the Library, IT Department, etc. as needed.)
3. Sample Syllabus

Department Chair Approval: Kenneth D. Machande Date: 10/31/12

CCC Chair Approval: Gail Brooks Date: 11/7/12

UCC Chair Approval: Date: 

New Course Proposal Cover Sheet (July 2012)
Rationale Statement:

This new course is intended to be part of the new minor program proposed to begin in the fall of 2013 (Data Science). Data Science is inherently computational, and several critical abilities are needed to work in this discipline. These include the ability to perform sophisticated queries and data manipulation using programming and specialized query languages, as well as the ability to construct analytics-based applications that make use of existing software libraries and algorithm packages for tasks such as classification, optimization, regression, and data fusion. Additionally, it is necessary for students to be able to develop applications that work on extremely large and varied data sets. This course is intended to provide students extensive training and practical experience in the manipulation of data and in the development of sophisticated Data Science applications.

Impact Statement (Provide details about the Library, space, budget, and technology impacts created by adding this new course. Include supporting statements from the Library, IT Department, etc. as needed.):

The impact of this specific course will be minimal. It is anticipated that current faculty members will teach the course as part of their regular teaching load. No new library resources are needed, and all required software will be open-source and freely available.
BUAD 400: Analytics Application Development

Instructor: Christopher J. Garcia, Ph.D.
Office: George Washington Hall 215-B
Telephone: (540) 654-1456
Email: cgarcia@umw.edu
Office hours: TR 3:00 p.m.-5:00 p.m., F 4:00-5:00 p.m. and by appointment
Prerequisite: Grade of C or better in CPSC 220 or equivalent
NOTE: This is a programming-intensive course

Course Description:
A course in programming and data manipulation techniques for constructing analytics-based applications. Topics include SQL and no-SQL databases, using web service API’s to acquire data, introduction to Hadoop and MapReduce, and use of third-party analytic component API's.

Learning Outcomes:
1) Students are able to use modern programming languages and query languages to clean, transform, and manipulate data.
2) Students are able to use third-party analytics API’s and web service interfaces
3) Students are able to develop Analytics-based applications using a modern programming language and software stack

Required Textbooks:
Programming Collective Intelligence (2007), Toby Segaran, O’Reilly:Sebastopol, CA
MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems (2012), Miner and Shook, O’Reilly:Sebastopol, CA
Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites (2011), Matthew Russell, O’Reilly:Sebastopol, CA

Grading and Assigned Work:
Two programming assignments each counting 25% of final grade:
- **Assignment 1:** The first assignment will focus on large-scale data handling and summarization. This will entail mining through large volumes of messy web-based data and applying processing and summarization methods using the MapReduce paradigm.
- **Assignment 2:** The second assignment will focus on applying advanced algorithms and analytical models to big data to include classifiers, regression methods, and optimization. You will develop an application that makes sense of big data (for example, a sentiment analysis tool for analyzing blogs) or prescribes an optimal solution for a complex problem (e.g. vehicle fleet routing and scheduling).

One course project completed over the whole semester counting 50% of final grade:
You will work in groups of three or four to complete this project. A project proposal is due by the fourth week of the semester, and the final project will be due during the last week of class. The project should be of interest to you and your teammates and should involve at least one key analytic element at the heart of the problem. There will be several group meetings with the instructor throughout the semester to review progress and adjust scope if needed.
For final grades a total grade of 93.34 and above receives an “A”, 90.00 to 93.33 is an “A-“, 87.34 to 89.99 is a “B+”, 83.34 to 87.33 is a “B”, 80.00 to 83.33 is a “B-“, 77.34 to 79.99 is a “C+”, 73.34 to 77.33 is a “C”, 70.00 to 73.33 is a “C-“, 67.34 to 69.99 is a “D+”, 60.00 to 67.33 is a “D”, Below 60 is an “F”. For Pass/Fail grading, a grade equivalent to a C- or higher is required to pass.

Attendance and Late Work:
Class attendance is necessary to succeed in this course. However, there is no formal attendance policy. It is the student’s responsibility to stay current on class material. Late work will be penalized by one-half letter grade per day late unless prior arrangements are made with the instructor.

Academic Integrity:
The University of Mary Washington does not tolerate academic dishonesty in any form. Penalties for cheating on exams or any other assignments in this course may include course failure and suspension or expulsion from the university. It is understood that all material submitted will be pledged in accordance with the Honor Code of UMW.

Disability Accommodations:
Accommodations will be made as needed as coordinated through the Office of Disability Resources.

Tentative Schedule:

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
<th>Assignment Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction to Hadoop and MapReduce</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Introduction to NoSQL Databases</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Introduction to Hadoop and MapReduce Part II</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Overview of Python Programming Programming Data Transforms in Python</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Overview of UML and Software Design API Concepts and Web Service Interfaces</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Data Visualization API’s part I</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Data Visualization API’s part II</td>
<td>Assignment I</td>
</tr>
<tr>
<td>8</td>
<td>Classification Methods and API’s Part I</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Numerical Prediction Methods and API’s Part I</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Optimization Methods and API’s</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Text Mining Fundamentals</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Document Filtering and Categorizing</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Mining Social Media and Blogs Part I</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Mining Social Media and Blogs Part II</td>
<td>Assignment II</td>
</tr>
<tr>
<td>15</td>
<td>Classification Methods and API’s Part II</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Numerical Prediction Methods and API’s Part II</td>
<td>Course Project</td>
</tr>
</tbody>
</table>