

San José State University
College of Science/Department of Computer Science
CS 274, Topics in XML and Web Intelligence, Section 1, Spring, 2016

Course and Contact Information

| | |
|-------------------------|--|
| Instructor: | Dr. Teng Moh |
| Office Location: | MQH 411 |
| Telephone: | (408) (924-5147) |
| Email: | MyFirstName <dot> MyLastName <at> SJSU <dot> EDU |
| Office Hours: | MW 20:45 to 21:45 [<i>See University Policy S12-1 at http://www.sjsu.edu/senate/docs/S12-1.pdf for guidelines</i>] |
| Class Days/Time: | MW 19:30 to 20:45 |
| Classroom: | MQH 422 |
| Prerequisites: | CS 160 or instructor consent. Familiarity with XML is assumed. |

Course Description

XML: DTD, Schema, Namespace, XSLT, XPath, Xquery, Encryption, Signature in XML, applications in vertical industries; Semantic Web; RDF, RDFS, Ontology, Inferences; Web services, relevant tools for search, inference, data conversion in XML, Semantic web applications. This section focuses specifically on Semantic Web and Social Networks.

Learning Outcomes

Upon successful completion of this course, students will be able to:

1. SLO 1 *understand Finding Similar Items*
2. SLO 2 *understand Mining Data Streams*
3. SLO 3 *understand Link Analysis*
4. SLO 4 *understand Frequent Itemsets*
5. SLO 5 *understand Advertising on the Web*
6. SLO 6 *understand Recommendation Systems*
7. SLO 7 *understand Mining Social-Network Graphs*
8. SLO 8 *implement new tools for web intelligence based on the above techniques*

Course Learning Outcomes (CLO)

Upon successful completion of this course, students will be able to:

1. CLO 1 *understand Finding Similar Items*

2. CLO 2 *understand Mining Data Streams*
3. CLO 3 *understand Link Analysis*
4. CLO 4 *understand Frequent Itemsets*
5. CLO 5 *understand Advertising on the Web*
6. CLO 6 *understand Recommendation Systems*
7. CLO 7 *understand Mining Social-Network Graphs*
8. CLO 8 *implement new tools for web intelligence based on the above techniques*

Required Texts/Readings

Textbook

- *Jure Leskovec, Anand Rajaraman, Jeff Ullman, **Mining of Massive Datasets**, Cambridge University Press, 2nd ed., 2014, ISBN: 978-1107077232*

Other Readings [Optional]

- *Ian H. Witten, Eibe Frank, and Mark A. Hall, **Data Mining**, Morgan Kaufmann, 3rd ed., 2011, ISBN: 978-0123748560*

Other equipment / material requirement

- *Wireless laptop*

Course Requirements and Assignments

SJSU classes are designed such that in order to be successful, it is expected that students will spend a minimum of forty-five hours for each unit of credit (normally three hours per unit per week), including preparing for class, participating in course activities, completing assignments, and so on. More details about student workload can be found in [University Policy S12-3](http://www.sjsu.edu/senate/docs/S12-3.pdf) at <http://www.sjsu.edu/senate/docs/S12-3.pdf>.

Homework is due typewritten (include source code, but not executable files) by class starting time on the due date. Each assigned problem requires a solution and an explanation (or work) detailing how you arrived at your solution. Cite any outside sources used to solve a problem. When grading an assignment, I may ask for additional information. A subset of the assigned problems will typically be graded.

Refer the course website for latest information of homework assignments.

NOTE that [University policy F69-24](http://www.sjsu.edu/senate/docs/F69-24.pdf) at <http://www.sjsu.edu/senate/docs/F69-24.pdf> states that “Students should attend all meetings of their classes, not only because they are responsible for material discussed therein, but because active participation is frequently essential to insure maximum benefit for all members of the class. Attendance per se shall not be used as a criterion for grading.”

Grading Policy

Grading information:

- *I will determine letter grades for the course, including +/- grades based on*

| <i>Percentage</i> | <i>Grade</i> |
|---------------------|--------------|
| <i>92 and above</i> | <i>A</i> |
| <i>90 - 91</i> | <i>A-</i> |
| <i>88 - 89</i> | <i>B+</i> |
| <i>82 - 87</i> | <i>B</i> |
| <i>80 - 81</i> | <i>B-</i> |
| <i>78 - 79</i> | <i>C+</i> |
| <i>72 - 77</i> | <i>C</i> |
| <i>70 - 71</i> | <i>C-</i> |
| <i>60 - 69</i> | <i>D</i> |
| <i>59 and below</i> | <i>F</i> |

- *List of the percentage weight [or point value] assigned to various class assignments*
 - *Homework: 20%*
 - *Midterm (Wednesday, Mar. 23): 20%*
 - *Project: 30%*
 - *Final (Wednesday, May 18, 19:45 – 22:00): 30%*
- **No** *make-up exams will be given and no late homework will be accepted.*

Note that “All students have the right, within a reasonable time, to know their academic scores, to review their grade-dependent work, and to be provided with explanations for the determination of their course grades.” See [University Policy F13-1](http://www.sjsu.edu/senate/docs/F13-1.pdf) at <http://www.sjsu.edu/senate/docs/F13-1.pdf> for more details.

Classroom Protocol

- **Wireless laptop is required.** *Your laptop must remain closed (preferably in your backpack and, in any case, not on your desk) until I inform you that it is needed for a particular activity.*
- **Cheating** *will not be tolerated, but working together is encouraged*
- *Student must be respectful of the instructor and other students. For example, but not limited*
 - *Turn off cell phones*
- *To encourage participation from students, no recording is allowed.*

University Policies

General Expectations, Rights and Responsibilities of the Student

As members of the academic community, students accept both the rights and responsibilities incumbent upon all members of the institution. Students are encouraged to familiarize themselves with SJSU’s policies and practices pertaining to the procedures to follow if and when questions or concerns about a class arises. See [University Policy S90–5](http://www.sjsu.edu/senate/docs/S90-5.pdf) at <http://www.sjsu.edu/senate/docs/S90-5.pdf>. More detailed information on a variety of related topics is available in the [SJSU catalog](http://info.sjsu.edu/web-dbgen/narr/catalog/rec-12234.12506.html), at <http://info.sjsu.edu/web-dbgen/narr/catalog/rec-12234.12506.html>. In general, it is recommended that students begin by seeking clarification or discussing concerns with their instructor. If such conversation is not possible, or if it does not serve to address the issue, it is recommended that the student contact the Department Chair as a next step.

Dropping and Adding

Students are responsible for understanding the policies and procedures about add/drop, grade forgiveness, etc. Refer to the current semester’s [Catalog Policies](http://info.sjsu.edu/static/catalog/policies.html) section at <http://info.sjsu.edu/static/catalog/policies.html>.

Add/drop deadlines can be found on the current academic year calendars document on the [Academic Calendars webpage](http://www.sjsu.edu/provost/services/academic_calendars/) at http://www.sjsu.edu/provost/services/academic_calendars/. The [Late Drop Policy](http://www.sjsu.edu/aars/policies/latedrops/policy/) is available at <http://www.sjsu.edu/aars/policies/latedrops/policy/>. Students should be aware of the current deadlines and penalties for dropping classes.

Information about the latest changes and news is available at the [Advising Hub](http://www.sjsu.edu/advising/) at <http://www.sjsu.edu/advising/>.

Consent for Recording of Class and Public Sharing of Instructor Material

[University Policy S12-7](http://www.sjsu.edu/senate/docs/S12-7.pdf), <http://www.sjsu.edu/senate/docs/S12-7.pdf>, requires students to obtain instructor's permission to record the course and the following items to be included in the syllabus:

- “Common courtesy and professional behavior dictate that you notify someone when you are recording him/her. You must obtain the instructor's permission to make audio or video recordings in this class. Such permission allows the recordings to be used for your private, study purposes only. The recordings are the intellectual property of the instructor; you have not been given any rights to reproduce or distribute the material.”
 - It is suggested that the greensheet include the instructor's process for granting permission, whether in writing or orally and whether for the whole semester or on a class by class basis.
 - In classes where active participation of students or guests may be on the recording, permission of those students or guests should be obtained as well.
- “Course material developed by the instructor is the intellectual property of the instructor and cannot be shared publicly without his/her approval. You may not publicly share or upload instructor generated material for this course such as exam questions, lecture notes, or homework solutions without instructor consent.”

Academic integrity

Your commitment, as a student, to learning is evidenced by your enrollment at San Jose State University. The [University Academic Integrity Policy S07-2](http://www.sjsu.edu/senate/docs/S07-2.pdf) at <http://www.sjsu.edu/senate/docs/S07-2.pdf> requires you to be honest in all your academic course work. Faculty members are required to report all infractions to the office of Student Conduct and Ethical Development. The [Student Conduct and Ethical Development website](http://www.sjsu.edu/studentconduct/) is available at <http://www.sjsu.edu/studentconduct/>.

Campus Policy in Compliance with the American Disabilities Act

If you need course adaptations or accommodations because of a disability, or if you need to make special arrangements in case the building must be evacuated, please make an appointment with me as soon as possible, or see me during office hours. [Presidential Directive 97-03](http://www.sjsu.edu/president/docs/directives/PD_1997-03.pdf) at http://www.sjsu.edu/president/docs/directives/PD_1997-03.pdf requires that students with disabilities requesting accommodations must register with the [Accessible Education Center](http://www.sjsu.edu/aec) (AEC) at <http://www.sjsu.edu/aec> to establish a record of their disability.

CS 274, Spring 2016, Course Schedule

The schedule is subject to change with fair notice and the notice will be made available in class.

Course Schedule

| Week | Date | Topics, Readings, Assignments, Deadlines |
|------|------|--|
| 1 | 2/1 | Topics in Web Intelligence |
| 1 | 2/3 | Shingling |
| 2 | 2/8 | Min-Hashing |
| 2 | 2/10 | Locality-Sensitive Hashing |
| 3 | 2/15 | Sampling a Fixed-Size Sample |
| 3 | 2/17 | Queries Over a Sliding Window |
| 4 | 2/22 | DGIM Method |
| 4 | 2/24 | Web Search |
| 5 | 2/29 | PageRank |
| 5 | 3/2 | Block-Based Update Algorithm |
| 6 | 3/7 | Frequent Itemset Mining |
| 6 | 3/9 | Association Rules |
| 7 | 3/14 | A-Priori Algorithm |
| 7 | 3/16 | Multistage Algorithm |
| 8 | 3/21 | Review |
| 8 | 3/23 | Midterm Exam |
| 9 | 4/4 | Performance-Based Advertising |
| 9 | 4/6 | BALANCE Algorithm |
| 10 | 4/11 | Content-based Recommender Systems |
| 10 | 4/13 | Collaborative Filtering |
| 11 | 4/18 | Community Detection |
| 11 | 4/20 | Spectral Clustering |
| 12 | 4/25 | Applications of Mining Data Streams |
| 12 | 4/27 | Applications of Link Analysis |
| 13 | 5/2 | Applications of Frequent Itemsets |
| 13 | 5/4 | Applications of Advertising on the Web |
| 14 | 5/9 | Applications of Recommendation Systems |
| 14 | 5/11 | Applications of Mining Social-Network Graphs |

| Week | Date | Topics, Readings, Assignments, Deadlines |
|-------------|-------------|---|
| 15 | 5/16 | Review |
| Final Exam | | Wednesday, May18, 19:45 – 22:00 |