**CSE 573 SEMANTIC WEB MINING**

**Mid-Term Exam - SPRING 2021**

**400 Points (40% of course grade)**

**March 11h, 2021**

1. Take-home and open note exam.
2. This is an **individual work exam** and **you may not work in groups**.
3. The midterm is attached to these instructions.
4. Please note that this is an **individual exam** and you **may not work in groups**.

**Exam Submission Instructions**: You can (1) type your answers into this word document and convert it to PDF, or (2) write down your answers on plain papers, scan and upload them as a PDF at **Canvas > Assignments > Midterm Submissions - DUE: March 11, 2021 11:59pm** by mid-night tonight.

**Duration**: The exam should take only about 75 minutes but I am giving you plenty extra time to work on it, convert or scan it as a single PDF document and submit it at Canvas.

**Academic Integrity**: All students in this class are subject to ASU’s Academic Integrity Policy and you should acquaint yourselves with its content and requirements, including a strict prohibition against working together or copying from each other and plagiarism. All violations will be reported to the Dean’s office, who maintain records of all offenses.

I will be online at Skype if you have any questions. My SkypeID is hdavulcu. Or, you can call me at my cell phone 602.386.6565 if you have any questions.

**GRADING**

* The exam is worth 400 points.

**DO NOT FORGET TO WRITE YOUR NAME and ASUID on top of each exam sheet!**

***PART I: Semantic Web Languages* [100 points]**

**1. Translate the HTML information below into an XML object. [30 points]**



**2. Draw the graph representation for above XML object and its lower bound schema. [40 points]**

**3. Translate following XML-QL query with Regular Path Expressions to an English query. [30 points]**

**SELECT R.telephone**

**FROM Yelp.restaurant R**

**WHERE R.takeout=true**

***PART II: Frequent Item-Sets and Association Rules* [100 points]**

**4. Assume you have to apply the APRIORI algorithm assuming that the minimum support is 40% (4 out of 10) to the following set of 10 transactions that involve purchases of items A, B, C, D, E, F, G.**

**T1={A,B,C} T6={B, C, D, E}**

**T2={A,F,G} T7={E, F}**

**T3={B,C,G} T8={B, C, F, G}**

**T4={A, B, C, F} T9={B,C, F, G}**

**T5={B, C} T10={B, C, G}**

**Show how Apriori’s level-wise Item Set Generation algorithm works for the above example. Indicate what candidate itemsets will be generated in each pass, and which remain in the candidate set after pruning. [50 points]**

**5. Generate association rules with support 40% and confidence 100%. [50 points]**

***PART III: Clustering* [100 points]**

**6. Assume, you want to cluster 7 observations into 3 clusters using K-Means clustering algorithm. After first iteration clusters, C1, C2, C3 has following observations:  
  
C1: {(2,2), (4,4), (6,6)}  
  
C2: {(0,4), (4,0)}  
  
C3: {(5,5), (9,9)}  
  
What will be the cluster centroids if you want to proceed for the second iteration? [40 pts]**

**C1: (\_\_, \_\_); C2:(\_\_,\_\_), C3:(\_\_,\_\_)**

**7. What will be the clusters for the second iteration? [60 points]**

**C1: { }  
  
C2: { }   
  
C3: { }**

***PART IV: Web Ranking* [100 points]**

***8. A directed network N has vertices {A, B, C} and directed edges (A🡪 B), (B🡪 B), (B 🡪 C), (C 🡪 A).***

***i. Draw the network N and write down its column-normalized hyperlink matrix***

***M for the calculation of simplified PageRank. [30 points]***

***9. Is M an irreducible matrix? Explain why? [30 points]***

***10. Show the evolution of the initial page-rank vector x = <xA, xB, xC> = <1, 1, 1> until the simplified pagerank algorithm converges. [40 points]***

|  |  |  |  |
| --- | --- | --- | --- |
| ***iteration*** | ***XA*** | ***XB*** | ***XC*** |
| ***0*** |  |  |  |
| ***1*** |  |  |  |
| ***2*** |  |  |  |
| ***3*** |  |  |  |
| ***4*** |  |  |  |
| ***5*** |  |  |  |
| ***6*** |  |  |  |
| ***7*** |  |  |  |

**Hint.** Replace ***x*** by the product **M *x*** until page rankings converge.