

DS 320: Homework 4

Due: Tuesday, December 4, 2018, 11:59pm (EST)

Name: _____

Question	Points	Score
1	10	
2	10	
3	10	
4	10	
5	10	
Total	50	

1. (10 points) Write an HLRT Wrapper to extract the information (Speaker, Date) from the following HTML page.

Data Science Seminar Speakers (Fall 2018)

- Sharon Huang - *October 2*
- Naomi Altman - *November 3*
- Zihan Zhou - *December 4*

```
<html><body>
<h1>Data Science Seminar Speakers (Fall 2018) </h1>
<ul>
<li><b>Sharon Huang</b> - <i>October 2</i> <br>
<li><b>Naomi Altman</b> - <i>November 3</i> <br>
<li><b>Zihan Zhou</b> - <i>December 4</i> <br>
</ul>
</body></html>
```

2. (10 points) Compute universal and core solution instances for the following Data exchange setting and source instance:

Source S has one relation

$DeptEmp(dept_id, mgr_name, eid)$

The target schema T has two relations

$Dept(dpt_id, mgr_id, mgr_name)$

$Emp(eid, dpt_id)$

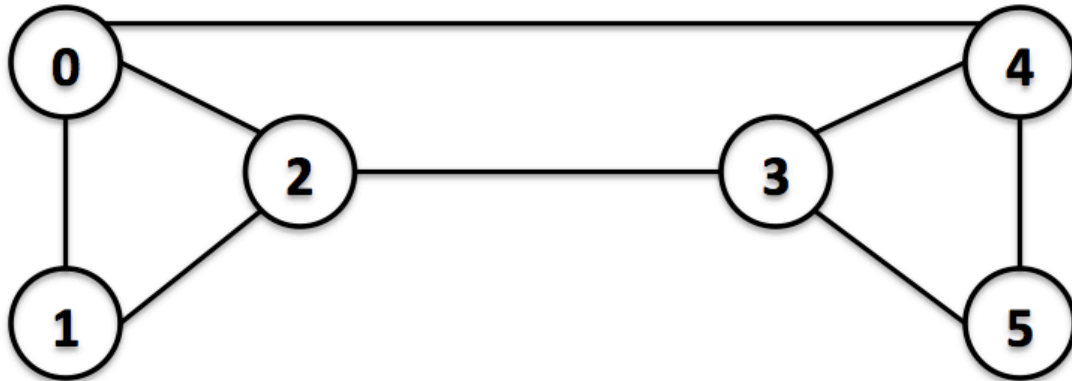
The mapping from S to T is:

$DeptEmp(d, n, e) \rightarrow \exists_M Dept(d, M, n), Emp(e, d)$

I(S) is $\{DeptEmp(IST, Liza, E005), DeptEmp(CSE, John, E010)\}$.

3. (10 points) what is the deep Web? What are the advantages and challenges of surfacing the deep Web?

4. (10 points) Cluster the following graph using SNN and a threshold equals one.



5. (10 points) Fuse the following two graphs using two different methods

