Papers:
(1) Privacy-Preserving Data Mining and the Need for Confluence of Research and Practice, and
(2) l-Diversity: Privacy Beyond k-Anonymity.

**Question 1** Name a few techniques used in privacy-preserving data mining. Apply some of the techniques to the following information:

<table>
<thead>
<tr>
<th>SSN</th>
<th>name</th>
<th>address</th>
<th>date-of-birth</th>
<th>condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>123456789</td>
<td>John Henry</td>
<td>25 Main St., A-Town, S-State, 44444</td>
<td>Jan. 1, 1945</td>
<td>High Blood Pressure</td>
</tr>
<tr>
<td>123123123</td>
<td>Mary Doe</td>
<td>1 First St., Shar, T-state, 55555</td>
<td>Feb. 8, 1961</td>
<td>Cancer</td>
</tr>
<tr>
<td>456456456</td>
<td>Joe Dibb</td>
<td>5 Fifth St., Rio, Y-state, 88888</td>
<td>May 11, 1933</td>
<td>Cancer</td>
</tr>
<tr>
<td>789789789</td>
<td>Jane Kara</td>
<td>7 Main, Shar, T-state, 55561</td>
<td>Sep. 6, 1952</td>
<td>High Blood Pressure</td>
</tr>
</tbody>
</table>

**Question 2** In your own words, explain what k-anonymity means. Give an example.

**Question 3** In your own words, explain what l-diversity means. Give an example.

**Paper: TIMBER: A native XML database**

**Question 4**

Explain in your own words how the start, end, and level labels can be used to determine ancestor-descendant relationship and parent-child relationship between two nodes.

**Question 5**

Consider an XML document with the following simplified DTD

```xml
<!ELEMENT bib (book* )>
<!ELEMENT book (title, author+, publisher, year, price )>
<!ELEMENT author (last, first )>
```

All undefined elements are #PCDATA.

Show how the query “List publishers of books written by Joe Doe” is specified in tree algebra.
Question 6
Consider a database $D$ containing one table $r(A, B, C)$. We have two views on $D$:
View $V_1(A, B)$ is defined as select $r.A, r.B$ from $r$.
View $V_2(B, C)$ is defined as select $r.B, r.C$ from $r$.
Assume $V_1$ contains tuples $(a_1, b_1)$ and $(a_1, b_2)$, and $V_2$ contains tuples $(b_1, c_1)$ and $(b_2, c_2)$.
What is the answer to the query select $r.A, r.C$ from $r$ given our knowledge about contents of $V_1$ and $V_2$? Make sure you consider closed world assumption as well as open world assumption.

Question 7
Consider the following relations

student (ID, name, major)
course (CNo, title, credits)
registration (ID, CNo)

consider the query

select student.ID, student.name, course.title
from student, course, registration
where student.ID = registration.ID and
course.CNo = registration.CNo and
course.credits > 3 and student.major = 'CS'

We have the following view

create view $V_1$ as
select course.CNo, course.title
from course, registration
where course.CNo = registration.CNo and
course.credits > 3

(a) Can $V_1$ be used for an equivalent rewriting of the query that does not use course and registration tables? If it can be used, then write the rewritten query. If it can not be used explain why.
(b) Can $V_1$ be used for an equivalent rewriting of the query (in this case, course or registration table can be used in the rewriting). If it can be used, then write the rewritten query. If it can not be used explain why.
(c) In both parts (a) and (b), if a rewriting is possible, explain whether the rewriting may be more efficient than the original query and why.

More questions related to information integration and DATALOG forthcoming.