Quiz 3 - Information Integration

1. What is the main problem with big data information integration?

- Pay-as-you-go model
- Probabilistic Schema Mapping
- **Many sources**
- Mediated Schema

2. What would be the two possible solutions associated with "big data" information integration as mentioned in lecture? (Choose 2)

- **Probabilistic Schema Mapping**
- Customer Transactions
- **Pay-as-you-go Model**
- Mediated Schema
- Attribute Grouping

3. What are mediated schemas?

- Schemas created from customer info.
- Schemas created entirely from attribute grouping.
- A type of probabilistic schema mapping.
- **Schema created from integrating two or more schemas.**

4. In attribute grouping, how would one evaluate if two attributes should go together? (Choose 2)

- **Probability of Two Attributes Co-occurring**
- Integrated Views
- **Similarity of Attributes**
- Customer Interaction
- Candidate Designs

5. What is a data item?

- Data found in a customer transaction.
- **Data that represents an aspect of a real-world entity.**
- The real worth of a data value.
- Data found in a mediated schema.

6. What is data fusion?

- Extracting a global value from a data source.
- Extracting true sources from a data source.
- **Extracting the true value of a data item.**
- Another term for customer analytics.

7. What is a potential problem of having too many data sources as mentioned in lecture?

- Too much data processing required for compression.
- **Too many data values.**
- Schema mapping becomes impossible.
- None, the problem is not a problem when using big data methodologies.

8. What do we mean when we say "the true value of a data item"?

- **Extrapolated data from a data item that represents the worth of that item.**
- Data created from statistical estimations.
- Another term for data fusion.

9. What is a potential method to deal with too many data sources as mentioned in lecture?

- **Compare and weigh each source by their trustworthiness.**
- Randomly select a sample of sources to represent the various data sources.
- None, the more the better.
- Take less samples per tick.