Assignment #3

1. [RDF] (40)

a. (20) Consider a set of natural language sentences collected from Web pages.

i. A human can like another human.

ii. A human can have a gender property of a male, female or non-binary.

iii. A human can be the father of another human.

iv. A human can be the mother of another human.

v. A human can be married to another human.

vi. A human can have a BirthYear property of type “xs:Year”.

vii. If a human is married to another, then they like each other.

viii. If a human is a mother or father, the human is a parent.

Write a RDF schema and give a graphical presentation to describe these relationships.

b. (20) Write an instance of the RDF schema in A that express the following with a complete information that can be inferred from the schema.

i. Mary is a woman and she is John’s wife.

ii. Sophie, Sandra and Susan are women.

iii. Mary and John has a son Frank.

iv. John was born in 1950.

v. Frank was born in 1980.

vi. Susan is John's daughter.

vii. Susan was born in 1978.

viii. Frank likes Sophie.

ix. Sandra likes Frank.

All other information for everyone else is unknown.

Write a RDF document and give a graphical presentation to describe these facts.

3. [Property Graph] (20)

Create a Labeled Property Graph to describe the project management data shown in Lecture 2-3. Describe the vertices and the edges along with their labels and properties. Draw the graph using a Graph database tool (Neo4j or other)

Note: Refer to the table included in the attachment with this assignment.

4. **[Approximate Query Processing]** (40) This question continues our discussion on using data synopsis for query processing based on data-driven approximation. You are given a vector of numbers: [127, 71, 87, 31, 59, 3, 43, 99, 100, 42, 0, 58, 30, 88, 72, 130], each data point records the frequency of communication of a server in a 5-minute interval. For example, in the first 5 minutes, 127 contacts are observed. In the next 5 minutes, 71 contacts, ...

Note: For these questions, similar to the examples shown in the lectures, discard the lowest level (high-resolution) coefficients (i.e. only keep the first 50% of coefficients).

a. (20) Give the Haar decomposition and draw a corresponding error tree for the contacts data vector.

b. (10) Give the process and result for reconstructing the frequency during time interval [15, 20] using
Haar decomposition

c. (10) Use Haar decomposition and error tree to compute the total number of communications between time interval [15, 30].