**Washington State University  
CPT\_S 415 – Big Data Online**

Srinivasulu Badri

**Assignment 3**

Name: Nam Jun Lee

Student Number: 11606459

**1. [RDF]**

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

1. A human can like another human.
2. A human can have a gender property of a male, female or non-binary.
3. A human can be the father of another human.
4. A human can be the mother of another human.
5. A human can be married to another human.
6. A human can have a Birth Year property of type “xs:Year”.
7. If a human is married to another, then they like each other.
8. 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.

RDF schema:

**<rdfs:**Classrdf**:**ID**=**"Human"**>**

**<rdfs:**comment**>**

The class of Human.

Superclass of Man, Woman, NonBinary, Year classes

**</rdfs:comment>**

**</rdfs:Class>**

**<rdfs:**Classrdf**:**ID**=**"AnotherHuman"**>**

**<rdfs:**comment**>**

The class of AnotherHuman

This class represents other human

**</rdfs:comment>**

**</rdfs:Class>**

**<rdfs:**Classrdf**:**ID**=**"Male"**>**

**<rdfs:**comment**>**

The class of Man

Subclass of Human

**</rdfs:comment>**

**<rdfs:**subClassOfrdf**:**resource**=**"#Human"**/>**

**</rdfs:Class>**

**<rdfs:**Classrdf**:**ID**=**"Female"**>**

**<rdfs:**comment**>**

The class of Female

Subclass of Human

**</rdfs:comment>**

**<rdfs:**subClassOfrdf**:**resource**=**"#Human"**/>**

**</rdfs:Class>**

**<rdfs:**Classrdf**:**ID**=**"NonBinary"**>**

**<rdfs:**comment**>**

The class of NonBinary

Subclass of Human

**</rdfs:comment>**

**<rdfs:**subClassOfrdf**:**resource**=**"#Human"**/>**

**</rdfs:Class>**

**<rdfs:**Classrdf**:**ID**=**"Year"**>**

**<rdfs:**comment**>**

The class of Year

Birth Year for Human

Subclass of Human

**</rdfs:comment>**

**<rdfs:**subClassOfrdf**:**resource**=**"#Human"**/>**

**</rdfs:Class>**

**<rdf:**Propertyrdf**:**ID**=**"canLike"**>**

**<rdfs:**comment**>**

Inherits its domain ("Human") and range ("AnotherHuman")

from its superproperty "isMarried"

**</rdfs:comment>**

**<rdfs:**domainrdfs**:**resource**=**"#Human" **/>**

**<rdfs:**rangerdfs**:**resource**=**"#AnotherHuman" **/>**

**<rdfs:**subPropertyOfrdfs**:**resource**=**"#isMarried" **/>**

**</rdf:Property>**

**<rdf:**Propertyrdf**:**ID**=**"canBe"**>**

**<rdfs:**domainrdfs**:**resource**=**"#Human" **/>**

**<rdfs:**rangerdfs**:**resource**=**"#Man" **/>**

**<rdfs:**rangerdfs**:**resource**=**"#Woman" **/>**

**<rdfs:**rangerdfs**:**resource**=**"#NonBinary" **/>**

**</rdf:Property>**

**<rdf:**Propertyrdf**:**ID**=**"isMarried"**>**

**<rdfs:**domainrdfs**:**resource**=**"#Human" **/>**

**<rdfs:**rangerdfs**:**resource**=**"#AnotherHuman" **/>**

**</rdf:Property>**

**<rdf:**Propertyrdf**:**ID**=**"isMotherOf"**>**

**<rdfs:**comment**>**

Inherits its domain ("Woman") and range ("AnotherHuman")

Inherits its domain ("NonBinary") and range ("AnotherHuman")

from its superproperty "isParent"

**</rdfs:comment>**

**<rdfs:**domainrdfs**:**resource**=**"#Woman" **/>**

**<rdfs:**domainrdfs**:**resource**=**"#NonBinary" **/>**

**<rdfs:**rangerdfs**:**resource**=**"#AnotherHuman" **/>**

**<rdfs:**subPropertyOfrdfs**:**resource**=**"#isParentOf" **/>**

**</rdf:Property>**

**<rdf:**Propertyrdf**:**ID**=**"isFatherOf"**>**

**<rdfs:**comment**>**

Inherits its domain ("Man") and range ("AnotherHuman")

Inherits its domain ("NonBinary") and range ("AnotherHuman")

from its superproperty "isParent"

**</rdfs:comment>**

**<rdfs:**domainrdfs**:**resource**=**"#Man" **/>**

**<rdfs:**domainrdfs**:**resource**=**"#NonBinary" **/>**

**<rdfs:**rangerdfs**:**resource**=**"#AnotherHuman" **/>**

**<rdfs:**subPropertyOfrdfs**:**resource**=**"#isParentOf" **/>**

**</rdf:Property>**

**<rdf:**Propertyrdf**:**ID**=**"isParentOf"**>**

**<rdfs:**domainrdfs**:**resource**=**"#Human" **/>**

**<rdfs:**rangerdfs**:**resource**=**"#AnotherHuman" **/>**

**</rdf:Property>**

**<rdf:**Propertyrdf**:**ID**=**"BirthYear"**>**

**<rdfs:**romainrdfs**:**resource**=**"#Human" **/>**

**<rdfs:**rangerdfs**:**resource**=**"#Year" **/>**

**</rdf:Property>**

Graphical Presentation of this schema:

NonBinary

Woman

Man

canBe

Human

Year

BirthYear

isMotherOf

isFatherOf

isMarried

canLike

isParentOf

Domain:

Range:

AnotherHuman

SubClassOf:

SubPropertyOf:

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

1. Mary is a woman and she is John’s wife.
2. Sophie, Sandra and Susan are women.
3. Mary and John has a son Frank.
4. John was born in 1950.
5. Frank was born in 1980.
6. Susan is John’s daughter.
7. Susan was born in 1978.
8. Frank likes Sophie
9. Sandra likes Frank.

All other information for everyone else is unknown.

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

RDF document:

**<rdf:**RDF

xmlns**:**rdf**=**"http://www.w3.org/1999/02/22-rdf-syntax-ns#"

xmlns**:**xsd**=**"http://www.w3.org/2001/XMLSchema#"

xmlns**:**uni**=**"http://www.mydomain.org/uni-ns"**>**

**<rdf:**Descriptionrdf**:**about**=**"Human1"**>**

**<uni:**name**>**Mary**</uni:name>**

**<uni:**gender**>**Woman**</uni:gender>**

**<uni:**BirthYear**></uni:BirthYear>**

**<uni:**Likesrdf**:**resource**=**"Human2"**/>**

**<uni:**isMotherOfrdf**:**resource**=**"Human3"**/>**

**<uni:**isParentOfrdf**:**resource**=**"Human3"**/>**

**</rdf:Description>**

**<rdf:**Descriptionrdf**:**about**=**"Human2"**>**

**<uni:**name**>**John**</uni:name>**

**<uni:**gender**>**Man**</uni:gender>**

**<uni:**BirthYear**>**1950**</uni:BirthYear>**

**<uni:**Likesrdf**:**resource**=**"Human1"**/>**

**<uni:**isMarriedrdf**:**resource**=**"Human1"**/>**

**<uni:**isFatherOf**>**

**<rdf:**Bag**>**

**<rdf:\_1** rdf**:**resource**=**"Human3"**/>**

**<rdf:\_2** rdf**:**resource**=**"Human4"**/>**

**</rdf:Bag>**

**</uni:isFatherOf>**

**<uni:**isParentOf**>**

**<rdf:**Bag**>**

**<rdf:\_1** rdf**:**resource**=**"Human3"**/>**

**<rdf:\_2** rdf**:**resource**=**"Human4"**/>**

**</rdf:Bag>**

**</uni:isParentOf>**

**</rdf:Description>**

**<rdf:**Descriptionrdf**:**about**=**"Human3"**>**

**<uni:**name**>**Frank**</uni:name>**

**<uni:**gender**>**Man**</uni:gender>**

**<uni:**BirthYear**>**1980**</uni:BirthYear>**

**<uni:**Likesrdf**:**resource**=**"Human5"**/>**

**</rdf:Description>**

**<rdf:**Descriptionrdf**:**about**=**"Human4"**>**

**<uni:**name**>**Susan**</uni:name>**

**<uni:**gender**>**Woman**</uni:gender>**

**<uni:**BirthYear**>**1978**</uni:BirthYear>**

**</rdf:Description>**

**<rdf:**Descriptionrdf**:**about**=**"Human5"**>**

**<uni:**name**>**Sophie**</uni:name>**

**<uni:**gender**>**Woman**</uni:gender>**

**<uni:**BirthYear**></uni:BirthYear>**

**</rdf:Description>**

**<rdf:**Descriptionrdf**:**about**=**"Human6"**>**

**<uni:**name**>**Sandra**</uni:name>**

**<uni:**gender**>**Woman**</uni:gender>**

**<uni:**BirthYear**></uni:BirthYear>**

**<uni:**Likesrdf**:**resource**=**"Human3"**/>**

**</rdf:Description>**

**</rdf:RDF>**

Graphical Presentation of the instance of schema:

Instance:

Man

Woman

NonBinary

canBe

Year

Human

BirthYear

isFatherOf

isMotherOf

isMarried

canLike

isParentOf

AnotherHuman

Likes

BirthYear

BirthYear

Mary

1950

isMarried

John

isParentOf

isMotherOf

BirthYear

Sophie

isParentOf

isFatherOf

isParentOf

isFatherOf

Likes

BirthYear

BirthYear

1980

Sandra

Likes

BirthYear

1978

Frank

Susan

**3. [Property Graph]**

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.

This data has a total of seven labels:

1. Project (4)

2. Employee (14)

3. Assign\_hours (21)

4. Job (9)

5. Participated\_In (21)

6. Worked\_for (21)

7. Job\_Is. (14)

1 to 4 labels are vertices, and 5 to 7 labels are edges.

The properties of Each label:

Project: PROJ\_NUM, PROJ\_NAME

Employee: EMP\_NUM, EMP\_NAME, EMP\_HIREDATE

Job: JOB\_CODE, JOB\_CLASS, CHR\_HOUR

Assign: PROJ\_NUM, EMP\_NUM, ASSIGN\_HOURS

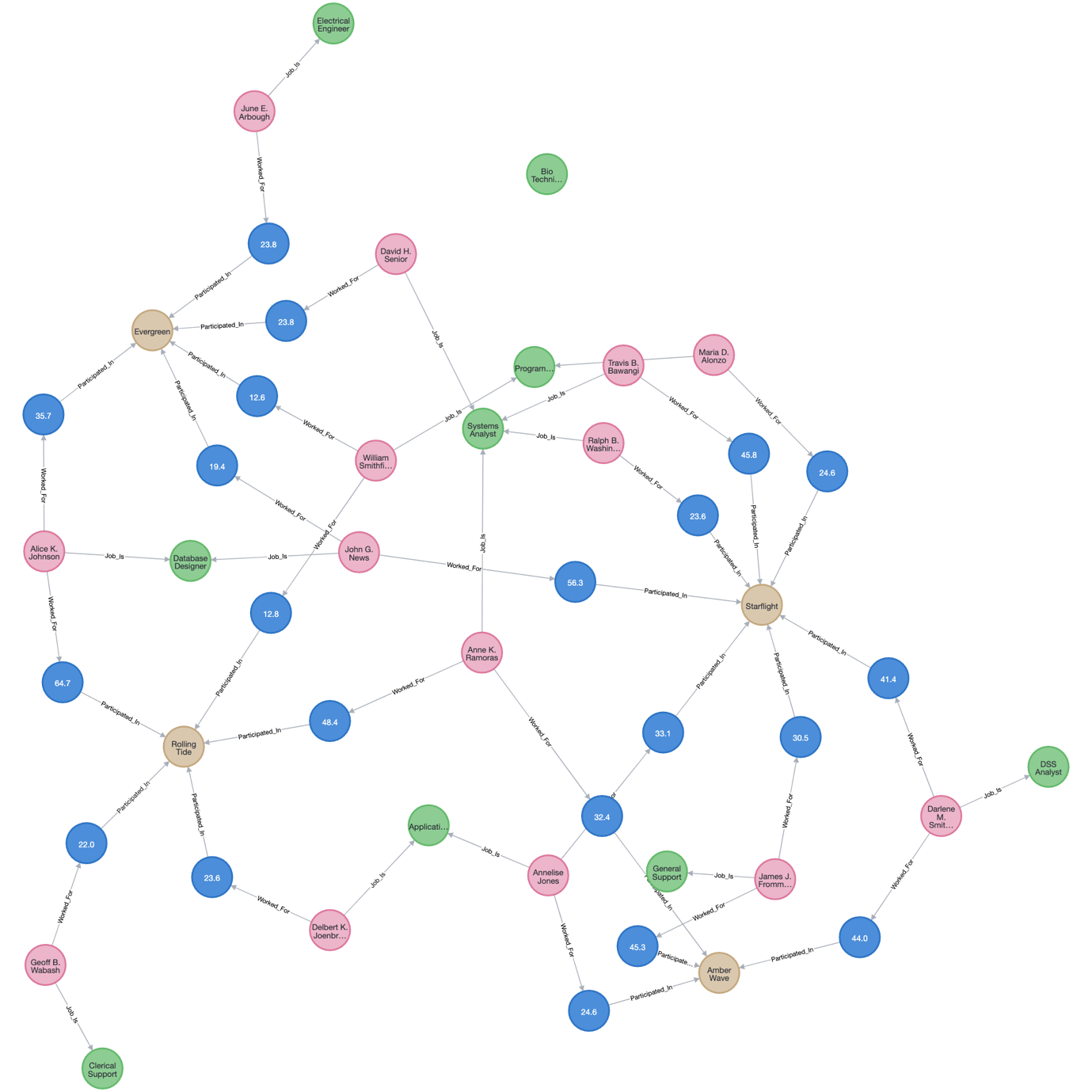
Connect Information:

Participated\_In edges connect Project vertices and Assign\_hours vertices.

Worked\_For edges connect Employee vertices and Assign\_hours vertices.

Job\_Is edges connect Employee vertices and Job vertices.

Graph using Neo4j:



4. **[Approximate Query Processing]**

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).

|  |  |  |
| --- | --- | --- |
| Resolution | Averages | Detail Coefficients |
| 4 | [127, 71, 87, 31, 59, 3, 43, 99, 100, 42, 0, 58, 30, 88, 72, 130] | ---- |
| 3 | [99, 59, 31, 71, 71, 29, 59, 101] | [28,28,28,-28,29,-29,-29,-29] |
| 2 | [79, 51, 50, 80] | [20,-21,21,-21] |
| 1 | [65, 65] | [14,-15] |
| 0 | [65] | [0] |

1. Give the Haar decomposition and draw a corresponding error tree for the contacts data vector.

According to table above,

Haar decomposition: [65, 0, 14, -15, 20, -21, 21, -21, 28, 28, 28, -28, 29, -29, -29, -29]

Error tree:

C0

65

C1

0

C3

C2

**+**

**-**

**-**

**-**

**+**

-15

14

C7

C6

C5

C4

**+**

**-**

**-**

**+**

-21

-20

21

20

C15

C14

C13

C12

C10

C8

C9

**-**

**-**

**+**

**+**

**+**

C11

**-**

-29

-29

-29

-28

28

28

28

29

**-**

**-**

**+**

**-**

**-**

**-**

**-**

**-**

**+**

**+**

**+**

**+**

**+**

**+**

**+**

127

71

87

59

3

43

99

100

42

0

72

130

88

30

58

31

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

C0

65

C1

0

C3

C2

**+**

**-**

**-**

**-**

**+**

-15

14

C7

C6

C5

C4

**+**

**-**

**-**

**+**

-21

-20

21

20

C15

C14

C13

C12

C10

C8

C9

**-**

**-**

**+**

**+**

**+**

C11

**-**

-29

-29

-29

-28

28

28

28

29

**-**

**-**

**+**

**-**

**-**

**-**

**-**

**-**

**+**

**+**

**+**

**+**

**+**

**+**

**+**

127

87

59

3

43

99

100

42

0

130

72

88

30

58

31

71

The frequency during time interval [15,20] is shown in red below. Where C0, C1, C2, C4, C9 are coefficients.

Therefore, A[15,20] = C0 + C1 + C2 – C4 – C9

= 65 + 0 + 14 – 20 – 28

= 31

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

C0

65

C1

0

C3

C2

**+**

**-**

**-**

**-**

**+**

-15

14

C7

C6

C5

C4

**+**

**-**

**-**

**+**

-21

-20

21

20

C15

C14

C13

C12

C10

C8

C9

**-**

**-**

**+**

**+**

**+**

C11

**-**

-29

-29

-29

-28

28

28

28

29

**-**

**-**

**+**

**-**

**-**

**-**

**-**

**-**

**+**

**+**

**+**

**+**

**+**

**+**

**+**

127

71

87

59

3

43

99

100

42

0

130

72

88

30

58

31

To calculate the total number of communications [15, 30] between time intervals, we must first find the time intervals [15, 20], time intervals [20, 25], and time intervals [25, 30].

The frequency during time interval [15,20] is shown in red below. Where C0, C1, C2, C4, C9 are coefficients.

Therefore, A[15,20] = C0 + C1 + C2 – C4 – C9

= 65 + 0 + 14 – 20 – 28

= 31

The frequency during time interval [20,25] is shown in red below. Where C0, C1, C2, C5, C10 are coefficients.

Therefore, A[20,25] = C0 + C1 - C2 + C5 + C9

= 65 + 0 - 14 + (-20) + 28

= 59

The frequency during time interval [25,30] is shown in red below. Where C0, C1, C2, C5, C10 are coefficients.

Therefore, A[25,30] = C0 + C1 - C2 + C5 - C9

= 65 + 0 - 14 + (-20) - 28

= 3

Hence, the total number of communications between time interval [15, 30],

A[15,30] = A[15,20] + A[20,25] + A[25,30]

= 31 + 59 + 3

= 93