DSCI 558: Building Knowledge Graphs

Quiz 14 (5 minutes)

**Question 1 (7.5 points):**

Choose the best type of similarity for semantic labeling in the following scenarios: Value Similarity, Histogram Similarity, Attribute name similarity, Distribution similarity, Value range similarity (numeric Jaccard)

1. 2 columns have the same semantic types: Histogram

|  |  |
| --- | --- |
| **Col1** | **Col2** |
| F | Female |
| F | Male |
| M | Female |
| F | Female |

1. 2 columns have the same semantic types: Attribute name

|  |  |
| --- | --- |
| **City** | **City** |
| Los Angeles | Tempe |
| San Francisco | Mesa |
| San Diego | Phoenix |
| San Jose |  |

1. 2 columns have different semantic types: Value

|  |  |
| --- | --- |
| **Col1** | **Col2** |
| Los Angeles | New York |
| San Francisco | California |
| San Diego | Utah |
| San Jose | Arizona |

1. 2 columns have different semantic types: Distribution

|  |  |
| --- | --- |
| **Col1** | **Col2** |
| 23 | 23 |
| 17 | 14 |
| 14 | 35 |
| 15 | 48 |

1. 2 columns have the same semantic types: Value range

|  |  |
| --- | --- |
| **Col1** | **Col2** |
| 35 | 23 |
| 17 | 14 |
| 14 | 35 |
| 15 | 30 |

**Question 2 (2.5 points):**

Select the **MINIMAL** set of existing sources that can use to explain s($location1,$location2,distance):

S1($location, lat)

S2($location, long)

S3($lat1, $long1, $lat2, $long2, distance) X

S4($location, zip)

S5($zip, long, lat)

S6($location, lat, long) X