Semantic Data Integration Report (Team 1)

Mohsen Safi Najafabadi, Mohammadamin Azin, Mohammad Almasi, Shabnam Hamzehloofard, Mona Manoochehri

Task3. Week 9 (22.06 – 07.07) - XML Parsing & Matching

(9th week)- Parse XML sources and create a schema tree for each source

Contributions:

Mohammad Almasi:
Write the latex report.
Maintained the git repository.
Dataset categorization and XML schema
Use `xml.etree.ElementTree` to convert excel to xml
Use `HtmlAgilityPack` tools to parse xml in C#
Implement XML parser to parse Schema files

Mona Manoochehri :
Write the latex report.
Use `xml.etree.ElementTree` to convert excel to xml
Use `HtmlAgilityPack` tools to parse xml in C#
Dataset categorization and XML schema
Implement XML parser to parse Schema files

Shabnam Hamzehloofard:
Write the latex report.
Use `xml.etree.ElementTree` to convert excel to xml
Use `HtmlAgilityPack` tools to parse xml in C#
Dataset categorization and XML schema
Implement XML parser to parse Schema files

Mohsen Safi Najafabadi:
Use `xml.etree.ElementTree` to convert excel to xml
Use `HtmlAgilityPack` tools to parse xml in C#
Dataset categorization and XML schema
Implement XML parser to parse Schema files

Mohammadamin Azin:
Use `xml.etree.ElementTree` to convert excel to xml
Dataset categorization and XML schema
Implement XML parser to parse Schema files

Semantic Data Integration Report (Team 1)

Mohsen Safi Najafabadi, Mohammadamin Azin, Mohammad Almasi, Shabnam Hamzehloofard, Mona Manoochehri

Task3. Week 10 (22.06 – 07.07) - XML Parsing & Matching

(9th week)- Parse XML sources and create a schema tree for each source

Contributions:

Mohammad Almasi:
Write the latex report.
Maintained the git repository.
Dataset categorization and XML schema
Parsing XML by Node
Parsing XML by Path
Implement XML by Path in Python
Implement XML by Node in Python
Implement XML by Path in C#
Implement XML by Node in C#
Implement `jaccard_similarity (token base)`

Mona Manoochehri :
Dataset categorization and XML schema
Parsing XML by Node
Parsing XML by Path
Implement XML by Path in Python
Implement XML by Node in Python

Shabnam Hamzehloofard:
Dataset categorization and XML schema
Parsing XML by Node
Parsing XML by Path
Implement XML by Path in Python
Implement XML by Node in Python

Mohsen Safi Najafabadi:
Dataset categorization and XML schema
Parsing XML by Node
Parsing XML by Path
Implement XML by Path in Python
Implement XML by Node in Python

Mohammadamin Azin:
Dataset categorization and XML schema
Parsing XML by Node
Parsing XML by Path
Implement XML by Path in Python
Implement XML by Node in Python

Evaluation – Path Similarity

CollegesUniversities XML:

```
<Root>
<Parrent>
    <Country>                                                    <!-- Country -->
        <Children>
            <Name>passau</Name>                                  <!-- Country/Name -->
            <latitude>125217.2395</latitude>                     <!-- Country/latitude -->
            <Longitude>34.75930829</Longitude>                   <!-- Country/Longitude -->
            <EstablishDate>2013-11-04T00</EstablishDate>         <!-- Country/EstablishDate -->
            <OtheInfo>http://nces.ed.gov.asp?ID=107840</OtheInfo> <!-- Country/OtheInfo -->
            <County>Pulaski</County>                             <!-- Country/County -->
        </Children>
    </Country>
</Parrent>

<Parrent>
    <Employee>                                                   <!-- Employee -->
        <Children>
            <HasDormitory>2</HasDormitory>                       <!-- Employee/HasDormitory -->
            <Address>604 Locust St</Address>                     <!-- Employee/Address -->
            <TotalEnrollment>52</TotalEnrollment>                <!-- Employee/TotalEnrollment -->
            <CapacityDormitory>0</CapacityDormitory>             <!-- Employee/CapacityDormitory -->
            <City>                                               <!-- Employee/City -->
                <Children>
                <ZIPCode>72114</ZIPCode>                         <!-- Employee/City/ZIPCode -->
                <Address2>NOT AVAILABLE</Address2>               <!-- Employee/City/Address2 -->
                <StudentCount>70</StudentCount>                  <!-- Employee/City/StudentCount -->
                <Telephone>(501) 374-6305 ext 107</Telephone>    <!-- Employee/City/Telephone -->
                <Website>www.shortercollege.org/</Website>       <!-- Employee/City/Website -->
                </Children>
            </City>
        </Children>
    </Employee>
</Parrent>
</Root>
```

CollegesUniversities(DataSet One) Path:

1. Country
2. Country/Name
3. Country/latitude
4. Country/Longitude
5. Country/EstablishDate
6. Country/OtheInfo
7. Country/County
8. Employee
9. Employee/HasDormitory
10. Employee/Address
11. Employee/TotalEnrollment
12. Employee/CapacityDormitory
13. Employee/City
14. Employee/City/ZIPCode
15. Employee/City/Address2
16. Employee/City/StudentCount
17. Employee/City/Telephone
18. Employee/City/Website

CollegeUniversityCampuses XML:

```xml
<Root>
<Parrent>
    <UniversityName>                                        <!-- UniversityName -->
        <Children>
            <Url>g</Url>                                    <!-- UniversityName/Url -->
            <AverageGPA>g</AverageGPA>                      <!-- UniversityName/AverageGPA -->
            <ContactName>g</ContactName>                    <!-- UniversityName/ContactName -->
            <EmployeeCount>g</EmployeeCount>                <!-- UniversityName/EmployeeCount -->
            <CountStudentWorkInUni>g</CountStudentWorkInUni>        <!-- UniversityName/CountStudentWorkInUni -->
            <NumberOfDisabledStudent>g</NumberOfDisabledStudent>    <!-- UniversityName/NumberOfDisabledStudent -->
            <AnnualScholarShipGranted>g</AnnualScholarShipGranted>  <!-- UniversityName/AnnualScholarShipGranted -->
            <NumberOfGraduatedStudents>g</NumberOfGraduatedStudents> <!-- UniversityName/NumberOfGraduatedStudents -->
            <County>                                        <!-- UniversityName/County -->
                <Children>
                    <ZIP>g</ZIP>                            <!-- UniversityName/County/ZIP -->
                    <Address>g</Address>                    <!-- UniversityName/County/Address -->
                </Children>
            </County>
        </Children>
    </UniversityName>
</Parrent>
</Root>
```

CollegeUniversityCampuses (DataSet Two) Path:

1. UniversityName
2. UniversityName/Url
3. UniversityName/AverageGPA
4. UniversityName/ContactName
5. UniversityName/EmployeeCount
6. UniversityName/CountStudentWorkInUni
7. UniversityName/NumberOfDisabledStudent
8. UniversityName/AnnualScholarShipGranted
9. UniversityName/NumberOfGraduatedStudents
10. UniversityName/County
11. UniversityName/County/ZIP
12. UniversityName/County/Address

NationalUniversitiesRankings XML:

```xml
<Root>
<Parrent>
    <Country>                                          <!-- Country -->
        <Children>
            <City>Borojerd</City>                      <!-- Country/City -->
            <ContactName>trh</ContactName>             <!-- Country/ContactName -->
            <Location>                                 <!-- Country/Location -->
                <Children>
                    <Zip>604 Locust St</Zip>           <!-- Country/Location/Zip -->
                    <Description>604 Locust St</Description>  <!-- Country/Location/Description -->
                </Children>
            </Location>
            <UndergradEnrollment>                      <!-- Country/UndergradEnrollment -->
                <Children>
                    <Date>245633</Date>                <!-- Country/UndergradEnrollment/Date -->
                    <Uni>245633</Uni>                  <!-- Country/UndergradEnrollment/Uni -->
                    <Rank>245633</Rank>                <!-- Country/UndergradEnrollment/Rank -->
                    <TuitionFees>245633</TuitionFees>  <!-- Country/UndergradEnrollment/TuitionFees -->
                </Children>
            </UndergradEnrollment>
        </Children>
    </Country>
</Parrent>
</Root>

```

NationalUniversitiesRankings(DataSet Three) Path:

1. Country
2. Country/City
3. Country/ContactName
4. Country/Location
5. Country/Location/Zip
6. Country/Location/Description
7. Country/UndergradEnrollment
8. Country/UndergradEnrollment/Date
9. Country/UndergradEnrollment/Uni
10. Country/UndergradEnrollment/Rank
11. Country/UndergradEnrollment/TuitionFees

MediatedSchema XML:

```xml
<Root>
<Parrent>
    <UniversityName>                                <!-- UniversityName -->
        <Children>
            <Name>g</Name>                          <!-- UniversityName/Name -->
            <Url>g</Url>                            <!-- UniversityName/Url -->
            <Uni>g</Uni>                            <!-- UniversityName/Uni -->
            <Location>g</Location>                  <!-- UniversityName/Location -->
            <TotalEmployee>g</TotalEmployee>        <!-- UniversityName/TotalEmployee -->
            <EmployeeCount>g</EmployeeCount>        <!-- UniversityName/EmployeeCount -->
            <Country>                               <!-- UniversityName/Country -->
                <Children>
                    <City>eghe</City>               <!-- UniversityName/Country/City -->
                    <Address>604 Locust St</Address>    <!-- UniversityName/Country/Address -->
                    <ZIPCode>604 Locust St</ZIPCode>    <!-- UniversityName/Country/ZIPCode -->
                    <County>gdfsd</County>          <!-- UniversityName/Country/County -->
                    <Website>604 Locust St</Website>    <!-- UniversityName/Country/Website -->
                </Children>
            </Country>
        </Children>
    </UniversityName>
</Parrent>
</Root>
```

MediatedSchema

1. UniversityName
2. UniversityName/Name
3. UniversityName/Url
4. UniversityName/Uni
5. UniversityName/Location
6. UniversityName/TotalEmployee
7. UniversityName/EmployeeCount
8. UniversityName/Country
9. UniversityName/Country/City
10. UniversityName/Country/Address
11. UniversityName/Country/ZIPCode
12. UniversityName/Country/County
13. UniversityName/Country/Website

Ground_truth_matrix (datasetOne):
This positions (`"1,8", "2,2", "7,12", "10,10", "13,9", "14,11", "18,13"`) must be true.

DatasetOne_meadiated schema(jarowinkler_similarity):

DS/MS	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.536	0.511	0.515	0.515	0.498	0.488	0.488	0.501	0.49	0.485	0.485	0.486	0.485
2	0.404	0.467	0.376	0.676	0.343	0.216	0.216	0.446	0.422	0.409	0.409	0.413	0.409
3	0.551	0.315	0.523	0.523	0.54	0.527	0.527	0.593	0.53	0.564	0.564	0.523	0.519
4	0.483	0.451	0.457	0.513	0.541	0.469	0.469	0.594	0.53	0.563	0.563	0.524	0.52
5	0.526	0.434	0.544	0.532	0.546	0.566	0.528	0.51	0.52	0.509	0.47	0.513	0.556
6	0.49	0.452	0.458	0.514	0.54	0.47	0.47	0.546	0.52	0.51	0.519	0.513	0.51
7	0.44	0.473	0.478	0.532	0.554	0.577	0.536	0.732	0.70	0.687	0.687	0.691	0.687
8	0.419	0.4	0.403	0.403	0.391	0.468	0.464	0.393	0.47	0.464	0.464	0.466	0.464
9	0.542	0.534	0.59	0.544	0.559	0.579	0.556	0.608	0.57	0.556	0.568	0.574	0.556
10	0.345	0.459	0.403	0.403	0.379	0.48	0.429	0.383	0.43	0.595	0.422	0.425	0.519
11	0.454	0.414	0.477	0.477	0.535	0.615	0.6	0.501	0.47	0.504	0.517	0.509	0.504
12	0.447	0.504	0.467	0.467	0.573	0.549	0.537	0.543	0.57	0.537	0.57	0.542	0.537
13	0.447	0.481	0.487	0.487	0.463	0.523	0.509	0.535	0.58	0.568	0.568	0.571	0.568
14	0.398	0.472	0.429	0.429	0.451	0.487	0.487	0.464	0.56	0.582	0.745	0.587	0.546
15	0.395	0.467	0.474	0.424	0.4	0.491	0.479	0.458	0.49	0.658	0.572	0.526	0.572
16	0.386	0.451	0.41	0.459	0.482	0.463	0.513	0.543	0.59	0.57	0.616	0.637	0.57
17	0.392	0.462	0.469	0.469	0.482	0.574	0.56	0.496	0.48	0.512	0.549	0.517	0.549
18	0.398	0.472	0.429	0.479	0.451	0.579	0.541	0.464	0.56	0.583	0.583	0.55	0.671

Cardinality_matrix(datasetOne):

DS/MS	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0	0	0	0	0	0	0	1	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	1	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	1	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	1	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	1	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	1

TP = 6, FP = 1, FN = 0

Precision = 6/6+1= 0.857
Recall = 1
F1-measure = 2*(0.857*1)/ 0.857+1= 0.922

Ground_truth_matrix (datasetTwo):

This positions (`"1,1", "2,3", "5,7", "10,12", "12,10"`) must be true.

DatasetTwo_meadiated schema(jarowinkler_similarity):

DS/MS	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.953	0.947	0.956	0.956	0.922	0.9	0.9	0.927	0.904	0.893	0.893	0.897	0.893
2	0.956	0.925	0.901	0.956	0.897	0.892	0.892	0.923	0.896	0.884	0.884	0.888	0.884
3	0.912	0.903	0.906	0.887	0.867	0.861	0.857	0.873	0.847	0.853	0.853	0.838	0.853
4	0.908	0.936	0.882	0.901	0.908	0.897	0.892	0.919	0.902	0.887	0.887	0.892	0.892
5	0.9	0.9	0.892	0.892	0.89	0.917	0.843	0.912	0.877	0.877	0.883	0.882	0.883
6	0.878	0.857	0.867	0.867	0.875	0.831	0.852	0.908	0.889	0.872	0.89	0.895	0.89
7	0.874	0.889	0.878	0.862	0.855	0.841	0.846	0.862	0.847	0.848	0.831	0.846	0.866
8	0.872	0.85	0.864	0.876	0.871	0.839	0.843	0.864	0.844	0.85	0.832	0.833	0.832
9	0.87	0.884	0.858	0.842	0.847	0.823	0.836	0.857	0.838	0.863	0.826	0.831	0.855
10	0.933	0.901	0.91	0.93	0.917	0.9	0.921	0.991	0.956	0.94	0.94	0.945	0.94
11	0.912	0.878	0.887	0.906	0.889	0.873	0.889	0.959	0.939	0.923	0.967	0.928	0.923
12	0.897	0.879	0.888	0.888	0.87	0.867	0.882	0.943	0.92	0.98	0.923	0.909	0.931

Cardinality_matrix(datasetTwo):

DS/MS	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	1	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	1	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	1	0	0	0

TP = 5
FP = 0
FN = 0

Precision = 1
Recall = 1
F1-measure = 1

Ground_truth_matrix (datasetThree):

This positions (`"1,8", "2,9", "4,5", "9,1"`) must be true.

DatasetThree_meadiated schema(jarowinkler_similarity):

DS/MS	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.436	0.411	0.415	0.415	0.498	0.488	0.488	0.401	0.49	0.485	0.485	0.486	0.485
2	0.458	0.494	0.5	0.423	0.534	0.516	0.516	0.346	0.62	0.609	0.609	0.613	0.609
3	0.581	0.584	0.547	0.461	0.622	0.548	0.536	0.323	0.609	0.578	0.578	0.597	0.592
4	0.49	0.507	0.514	0.465	0.664	0.561	0.561	0.426	0.532	0.519	0.564	0.568	0.519
5	0.465	0.478	0.484	0.476	0.627	0.565	0.565	0.418	0.57	0.556	0.627	0.597	0.556
6	0.494	0.487	0.538	0.447	0.61	0.54	0.548	0.483	0.617	0.513	0.545	0.613	0.577
7	0.548	0.466	0.537	0.421	0.528	0.542	0.564	0.475	0.536	0.575	0.53	0.535	0.519
8	0.53	0.53	0.514	0.483	0.596	0.605	0.545	0.497	0.613	0.643	0.614	0.638	0.602
9	0.533	0.45	0.46	0.496	0.32	0.418	0.44	0.453	0.478	0.479	0.416	0.323	0.458
10	0.53	0.489	0.514	0.369	0.596	0.576	0.535	0.497	0.583	0.614	0.584	0.609	0.562
11	0.513	0.469	0.491	0.419	0.568	0.599	0.617	0.467	0.607	0.629	0.576	0.611	0.614

Cardinality_matrix(datasetThree):

DS/MS	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	1	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	1	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0

TP = 3
FP = 2
FN = 0

Precision = 0.6
Recall = 1
F1-measure = 0.75

Evaluation – Node Similarity

```
nodes_datasetOne(
['Country'],
['Country','Name'],
['Country','latitude'],
['Country','Longitude'],
['Country','EstablishDate'],
['Country','OtheInfo'],
['Country','County'],
['Employee'],
['Employee','HasDormitory'],
['Employee','Address'],
['Employee','TotalEnrollment'],
['Employee','CapacityDormitory'],
['Employee','City'],
['Employee','City','ZIPCode'],
['Employee','City','Address2'],
['Employee','City','StudentCount'],
['Employee','City','Telephone'],
['Employee','City','Website']);

nodes_datasetTwo(
['UniversityName'],
['UniversityName','Url'],
['UniversityName','AverageGPA'],
['UniversityName','ContactName'],
['UniversityName','EmployeeCount'],
['UniversityName','CountStudentWorkInUni'],
['UniversityName','NumberOfDisabledStudent'],
['UniversityName','AnnualScholarShipGranted'],
['UniversityName','NumberOfGraduatedStudents'],
['UniversityName','County'],
['UniversityName','County','ZIP'],
['UniversityName','County','Address']);

nodes_datasetThree(
['Country'],
['Country','City'],
['Country','ContactName'],
['Country','Location'],
['Country','Location','Zip'],
['Country','Location','Description'],
['Country','UndergradEnrollment'],
['Country','UndergradEnrollment','Date'],
['Country','UndergradEnrollment','Uni'],
['Country','UndergradEnrollment','Rank'],
['Country','UndergradEnrollment','TuitionFees']);

nodes_mediatedSchema(
['UniversityName'],
['UniversityName','Name'],
['UniversityName','Url'],
['UniversityName','Uni'],
['UniversityName','Location'],
['UniversityName','TotalEmployee'],
['UniversityName','EmployeeCount'],
['UniversityName','Country'],
['UniversityName','Country','City'],
['UniversityName','Country','Address'],
['UniversityName','Country','ZIPCode'],
['UniversityName','Country','County'],
['UniversityName','Country','Website'])
```

Ground_truth_matrix(datasetOne):
This positions (`"1,8", "2,2", "7,12", "10,10", "13,9", "14,11", "18,13"`) must be true.

DatasetOne_meadiated schema(jaccard_similarity(list1, list2)):

DS/MS	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.49	0	0	0	0	0	0	0.456	0.5	0.333	0.333	0.333	0.333
2	0.5	0.638	0.333	0	0	0	0	0	0.333	0.25	0.25	0.25	0.25
3	0.5	0	0	0	0	0	0	0	0.333	0.25	0.25	0.25	0.25
4	0.5	0	0	0	0	0	0	0	0.333	0.25	0.25	0.25	0.25
5	0.5	0	0	0	0	0	0	0	0.333	0.25	0.25	0.25	0.25
6	0.5	0	0	0	0	0	0	0	0.333	0.25	0.25	0.25	0.25
7	0.5	0	0	0	0	0	0	0	0.333	0.25	0.25	0.667	0.25
8	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0.6	0.25	0	0
11	0	0	0	0	0	0	0	0	0.68	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0.26	0.25	0	0	0
14	0	0	0	0	0	0	0	0	0	0.2	0.551	0.2	0
15	0	0	0	0	0	0	0	0	0	0.2	0	0	0
16	0	0	0	0	0	0	0	0	0	0.2	0	0	0
17	0	0	0	0	0	0	0	0	0	0.2	0	0	0
18	0	0	0	0	0	0	0	0	0	0.2	0	0	0.221

Cardinality_matrix(datasetOne):

DS/MS	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0	0	0	0	0	0	0	1	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	1	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	1	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	1	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0

TP = 5
FP = 2
FN = 0

Precision = 0.714 , Recall = 1 , F1-measure = 0.833

Ground_truth_matrix(datasetTwo):

This positions (`"1,1", "2,3", "5,7", "10,12", "12,10"`) must be true.

DatasetTwo_meadiated schema(jaccard_similarity(list1, list2)):

DS/MS	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.333	0.5	0.5	0.5	0.5	0.5	0.333	0.333	0.333	0.333	0.333	0.333	0.333
2	0.5	0.333	0.512	0.333	0.333	0.333	0.333	0.25	0.25	0.25	0.25	0.25	0.25
3	0.5	0.333	0.333	0.333	0.333	0.333	0.333	0.25	0.25	0.25	0.25	0.25	0.25
4	0.5	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.25	0.25	0.25	0.25	0.25
5	0.5	0.333	0.333	0.333	0.333	0.333	0.655	0.333	0.25	0.25	0.25	0.25	0.25
6	0.5	0.333	0.333	0.333	0.333	0.333	0.333	0.25	0.25	0.25	0.25	0.25	0.25
7	0.333	0.333	0.333	0.333	0.333	0.333	0.25	0.25	0.25	0.25	0.25	0.25	0.25
8	0.5	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.25	0.25	0.25	0.25	0.25
9	0.5	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.25	0.25	0.25	0.25	0.25
10	0.468	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.25	0.25	0.25	0.25	0.25
11	0.5	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.25	0.25	0.25	0.367	0.25
12	0.333	0.25	0.25	0.25	0.25	0.25	0.25	0.2	0.2	0.2	0.2	0.5	0.2
13	0.333	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.2	0.5	0.2	0.5	0.2

Cardinality_matrix(datasetTwo):

DS/MS	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	1	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0.5	0	0	0	0

TP = 3
FP = 2
FN = 0

Precision = 0.6
Recall = 1
F1-measure = 0.75

Ground_truth_matrix(datasetThree):
This positions (`"1,8", "2,9","4,5", "9,4"`) must be true.

DatasetThree_meadiated schema(jaccard_similarity(list1, list2)):

DS/MS	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0	0	0	0	0	0	0	0.568	0.333	0.333	0.333	0.333	0.333
2	0	0	0	0	0	0	0	0.333	0.437	0.25	0.25	0.25	0.25
3	0	0	0	0	0	0	0	0.333	0.25	0.25	0.25	0.25	0.25
4	0	0	0	0	0.249	0	0	0.533	0.25	0.25	0.25	0.25	0.25
5	0	0	0	0	0.25	0	0	0.325	0.2	0.2	0.2	0.2	0.2
6	0	0	0	0	0.25	0	0	0.2	0.2	0.2	0.2	0.2	0.2
7	0	0	0	0	0	0	0	0.333	0.25	0.25	0.25	0.25	0.25
8	0	0	0	0	0	0	0	0.25	0.2	0.2	0.2	0.2	0.2
9	0	0	0	0.333	0	0	0	0.369	0.2	0.2	0.2	0.2	0.2
10	0	0	0	0	0	0	0	0.25	0.2	0.2	0.2	0.2	0.2
11	0	0	0	0	0	0	0	0.2	0.2	0.2	0.2	0.2	0.2
12	0.333	0	0	0	0	0	0	0	0.525	0.2	0.2	0.2	0.2

Cardinality_matrix(datasetThree):

DS/MS	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0	0	0	0	0	0	0	1	0	0	0	0	0
2	0	0	0	0	0	0	0	0	1	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	1	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0

TP = 3
FP = 1
FN = 0

Precision = 0.75
Recall = 1
F1-measure = 0.857