A Semantic Approach to Retrieving, Linking and Integrating Heterogeneous Geospatial Data

Ying Zhang
North China Electric Power University
dearzppzpp@gmail.com

Yao-Yi Chiang, Pedro Szekely, Craig A. Knoblock
Information Sciences Institute, University of Southern California
{yaoyichi, pszekely, knoblock}@isi.edu

The Problem

The ability of end-users to retrieve, combine and integrate geospatial data is limited.

- Different data structure
- Different information for the same entity
- Different geospatial data values (such as latitude, longitude, etc)

```
▼<place id="108462">
  <name>Lyon Recreational Center</name>
 ▼<url>
    http://wikimapia.org/108462/en/Lyon_Recreational_Center
  </url>
 ▼<location>
    <lon>-118.28833715</lon>
    <lat>34.0243829</lat>
    <north>34.0246941</north>
    <south>34.0240717</south>
    <east>-118.287853</east>
    <west>-118.2888213</west>
  </location>
 ▼<polygon>
    <point x="-118.2886389" y="34.0246941"/>
    <point x="-118.2888213" y="34.0244007"/>
    <point x="-118.288014" y="34.0240717"/>
    <point x="-118.287963" y="34.0240784"/>
    <point x="-118.2879227" y="34.0241518"/>
    <point x="-118.2879281" y="34.0242051"/>
    <point x="-118.287853" y="34.0243096"/>
    <point x="-118.2880193" y="34.0243874"/>
    <point x="-118.2880327" y="34.024374"/>
    <point x="-118.2881132" y="34.024414"/>
    <point x="-118.2880917" y="34.0244585"/>
    <point x="-118.2880917" y="34.0244585"/>
  </polygon>
</place>
```

From： Wikimapia

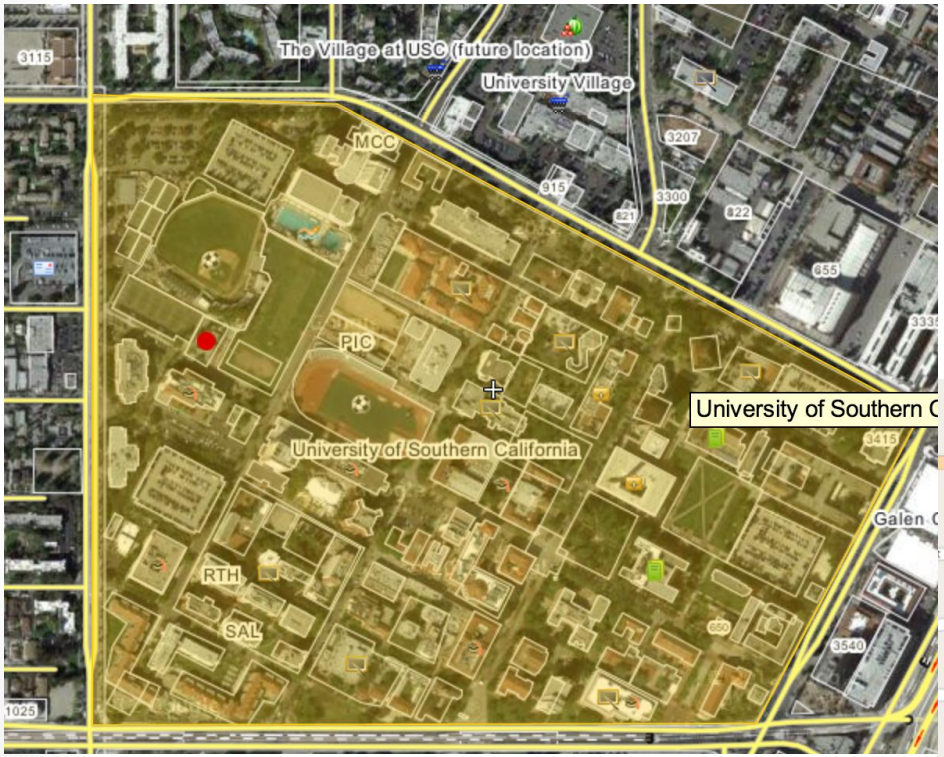
The Same Entity

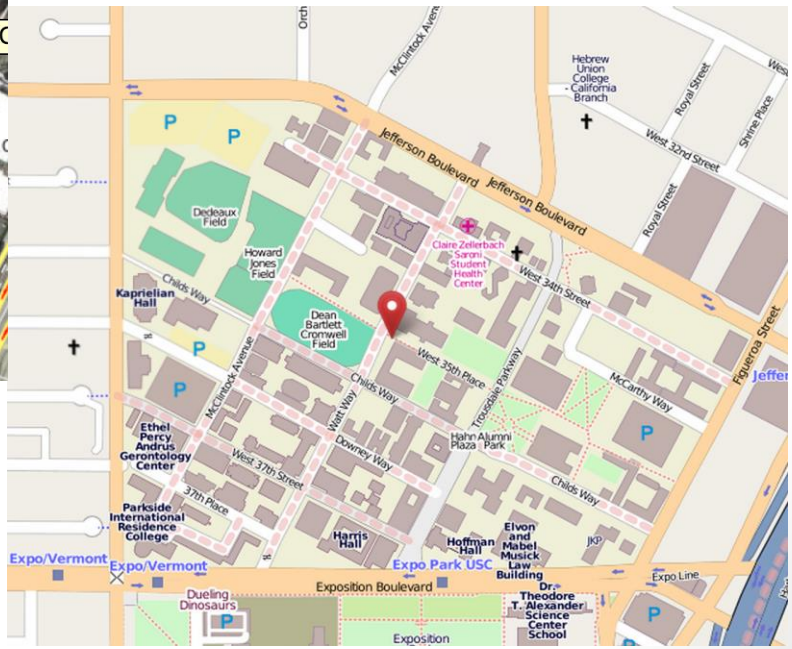
```
<node id="368166371" lat="34.0244572" lon="-118.2884081"
user="amillar" uid="28145" visible="true" version="1"
changeset="17182" timestamp="2009-04-01T07:19:47Z">
  <tag k="addr:state" v="CA"/>
  <tag k="building" v="yes"/>
  <tag k="ele" v="55"/>
  <tag k="gnis:county_name" v="Los Angeles"/>
  <tag k="gnis:feature_id" v="1657500"/>
  <tag k="gnis:import_uuid" v="57871b70-0100-4405-
bb30-88b2e001a944"/>
  <tag k="gnis:reviewed" v="no"/>
  <tag k="name" v="General William Lyon University Center"/>
  <tag k="source" v="USGS Geonames"/>
</node>
```

From： OpenStreetMap

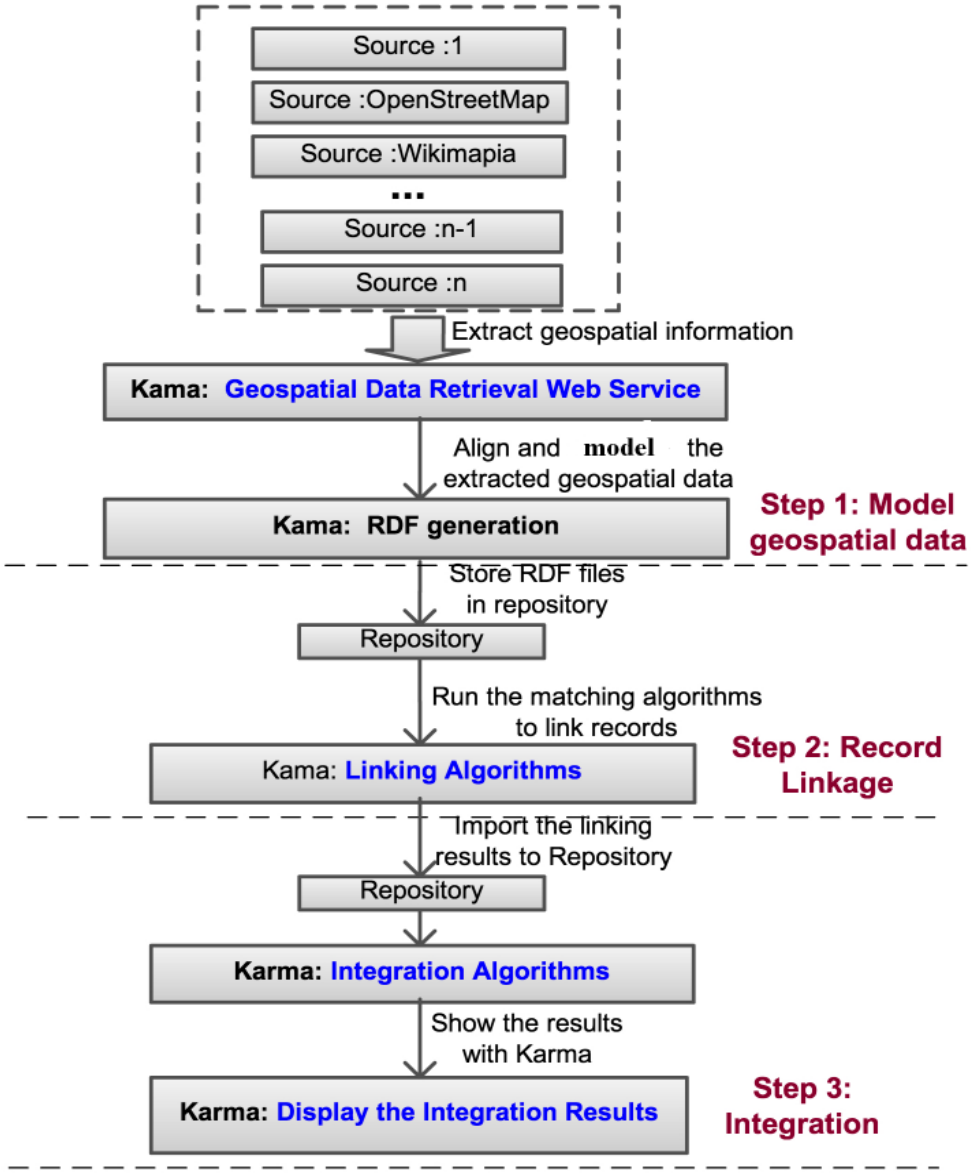
Motivation Example

- **Different sources have different information for the related entities**



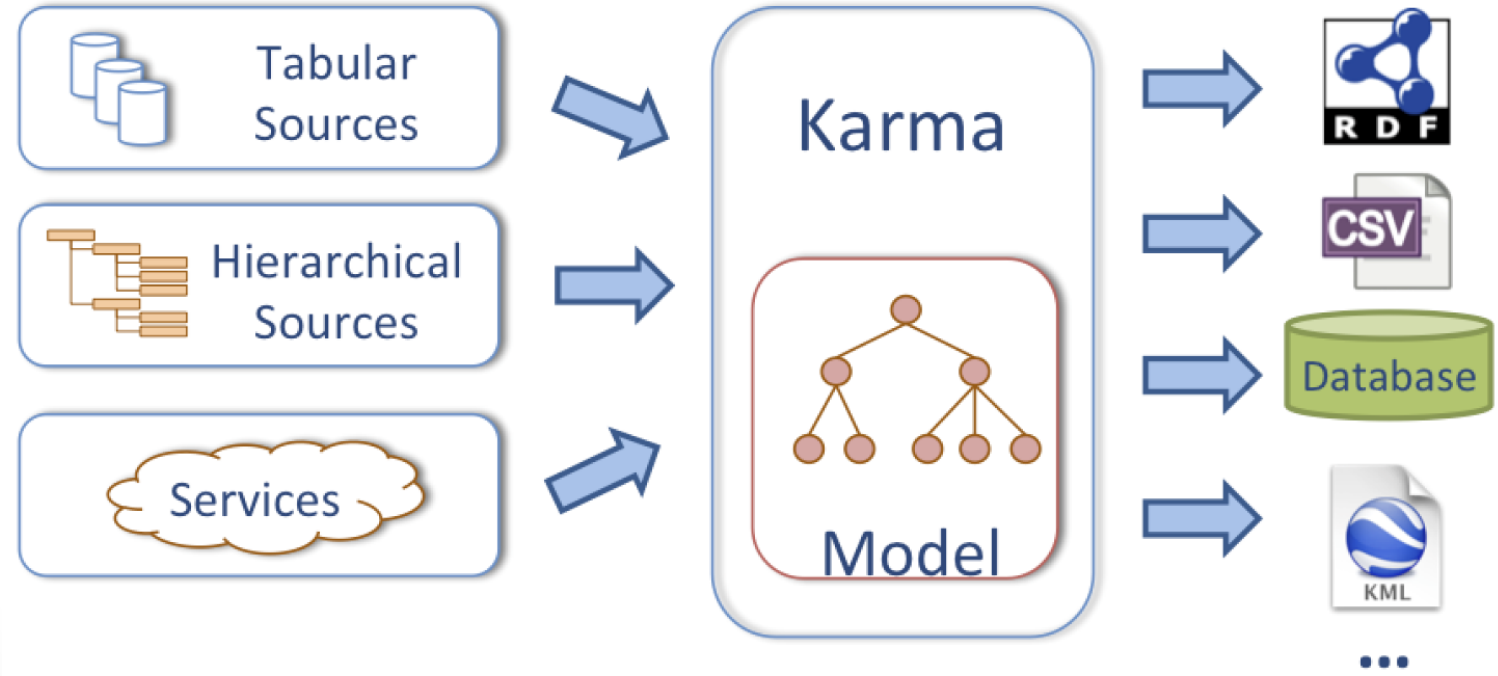


Overview


Source :1
Source :OpenStreetMap
Source :Wikimapia
...
Source :n-1
Source :n
Extract geospatial information
Kama: Geospatial Data Retrieval Web Service
Align and model the extracted geospatial data
Kama: RDF generation
Step 1: Model geospatial data
Store RDF files in repository
Repository
Run the matching algorithms to link records
Kama: Linking Algorithms
Step 2: Record Linkage
Import the linking results to Repository
Repository
Karma: Integration Algorithms
Show the results with Karma
Karma: Display the Integration Results
Step 3: Integration

Karma

Interactive tool for rapidly extracting, cleaning, transforming, integrating and publishing data



[Knoblock, Szekely, et al. Semi-automatically mapping structured sources into the semantic web. ISWC 2012]

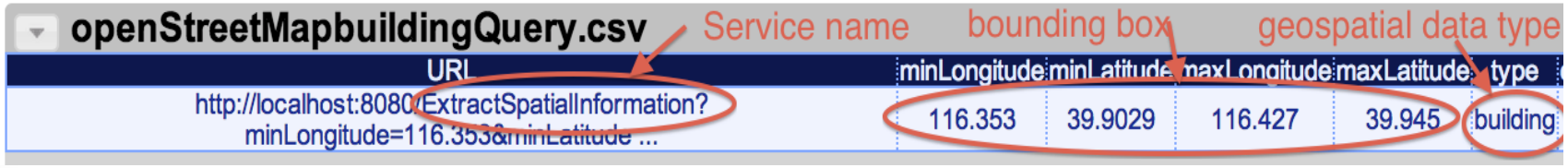
Outline

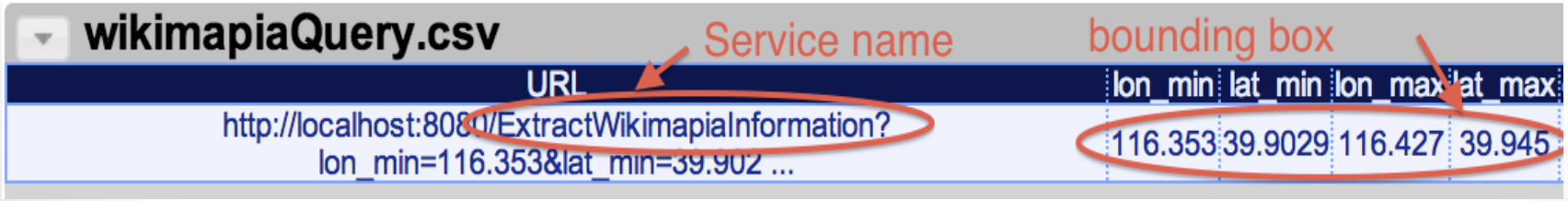
- ***Model the Geospatial Data***
- **Geospatial Data Linking**
- **Geospatial Data Integration**
- **Conclusion and Future work**

Model the Geospatial Data

Extract the geospatial data

- **Encapsulate the retrieval algorithms as Web services**
- **Embed all the inputs in a URL**

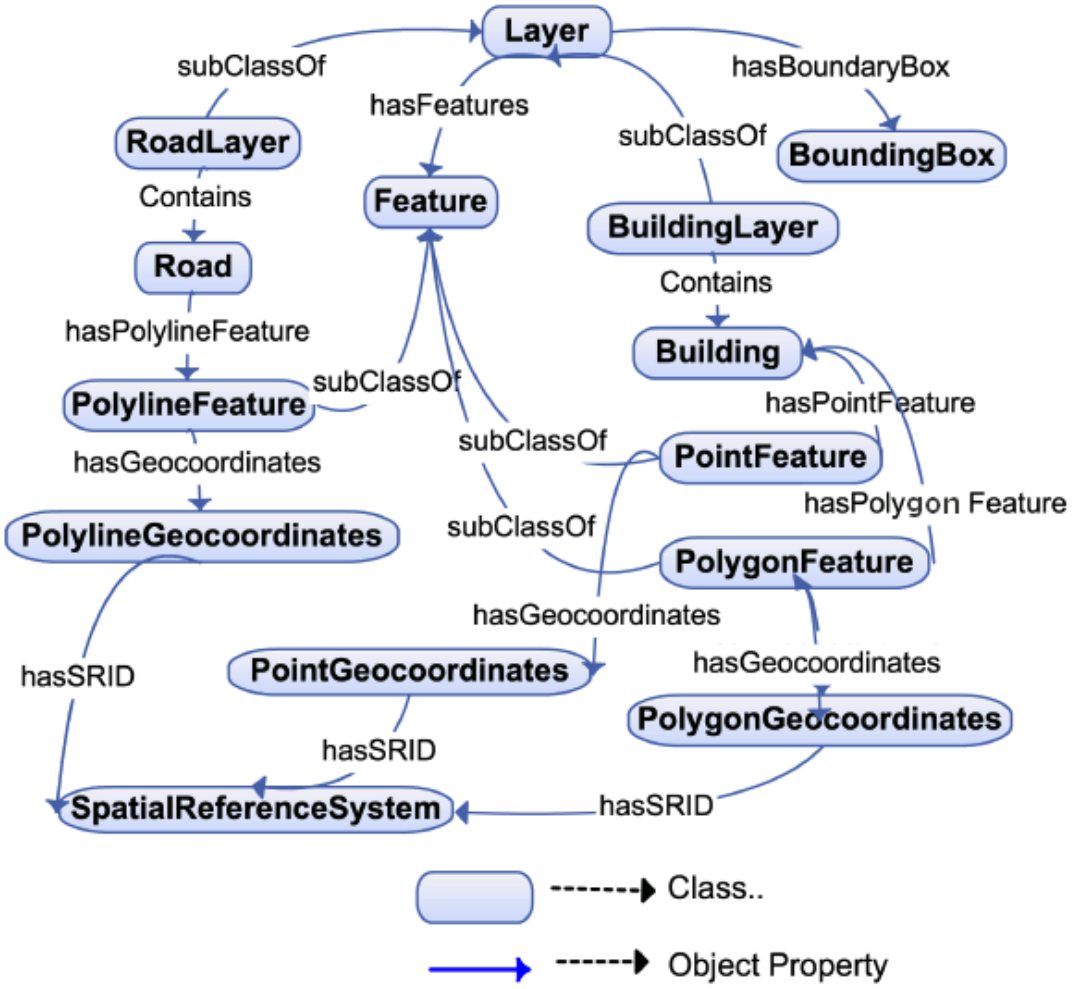




Model the Geospatial Data

Map the extracted geospatial data to RDF

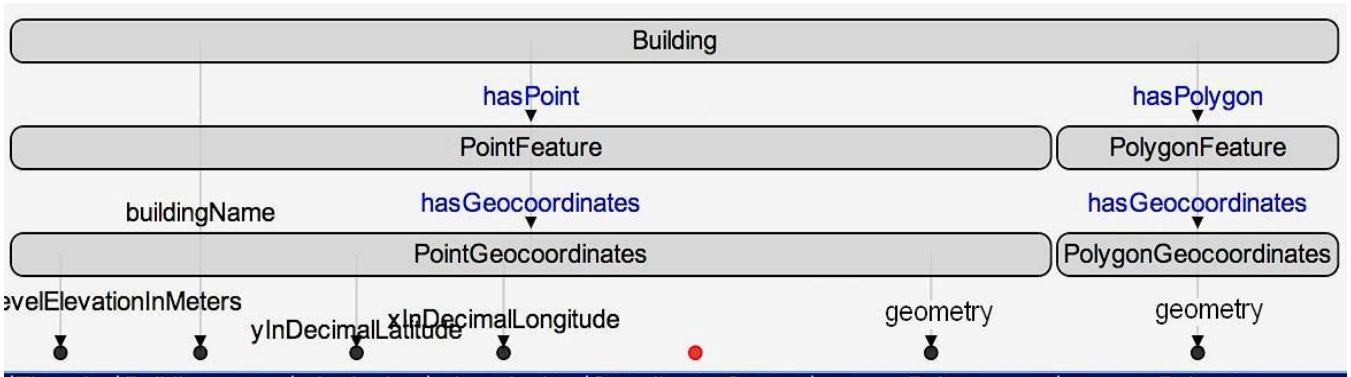
- **Build a generic geospatial ontology for aligning the extracted data**



Model the Geospatial Data

Map the extracted geospatial data to RDF

- **Model the extracted geospatial data by mapping them to RDF data with Karma**



Elevation	Building_name	Latitude	Longitude	Coordinate_System	Point	Polygon
56	United University Church	34.0230552	-118.2843721	WGS84	POINT(-118.2843721 34.0230552)	
56	Methodist Episcopal University Church (historical)	34.0232514	-118.2843209	WGS84	POINT(-118.2843209 34.0232514)	

Model the Geospatial Data

Map the extracted geospatial data to RDF

- **The generated N3 file**

building_wikimapia.n3

```
@prefix BuildingOntology: <http://www.semanticweb.org/ontologies/2012/9/BuildingOntology.owl#> .
@prefix s: <http://localhost:8080/source/wikimapia_building/> .

s:Building_1352239135666r1_0 a BuildingOntology:Building .
s:Point_1352239135666r1_1 a BuildingOntology:Point .
s:PointGeocoordinates_1352239135666r1_2 a BuildingOntology:PointGeocoordinates .
s:Polygon_1352239135666r1_3 a BuildingOntology:Polygon .
s:PolygonGeocoordinates_1352239135666r1_4 a BuildingOntology:PolygonGeocoordinates .
s:EPSGspecialReferenceSystemID_1352239135666r1_5 a BuildingOntology:EPSGspecialReferenceSystemID .
s:RelatedSourceNumer_58772 a BuildingOntology:RelatedSourceNumer .
s:Building_1352239135666r1_0 BuildingOntology:buildingName "New North Residential College" .
s:Building_1352239135666r1_0 BuildingOntology:hasPoint s:Point_1352239135666r1_1 .
s:Point_1352239135666r1_1 BuildingOntology:hasGeocoordinates s:PointGeocoordinates_1352239135666r1_2 .
s:PointGeocoordinates_1352239135666r1_2 BuildingOntology:xInDecimalLongitude "-118.2815335" .
s:PointGeocoordinates_1352239135666r1_2 BuildingOntology:yInDecimalLatitude "34.021055" .
s:Building_1352239135666r1_0 BuildingOntology:hasPolygon s:Polygon_1352239135666r1_3 .
s:Polygon_1352239135666r1_3 BuildingOntology:hasGeocoordinates s:PolygonGeocoordinates_1352239135666r1_4 .
s:PolygonGeocoordinates_1352239135666r1_4 BuildingOntology:hasSRID s:EPSGspecialReferenceSystemID_1352239135666r1_5 .
s:EPSGspecialReferenceSystemID_1352239135666r1_5 BuildingOntology:sridValue "4326" .
s:PolygonGeocoordinates_1352239135666r1_4 BuildingOntology:wellKnownBinary
"0103000020E6100000010000007000000A7052FFA0A925DC055DD239BAB024140BB438A0112925DC0B7D26BB3B1024140355EBA490C925DC0FBC
F7915DC06FF25B74B2024140910E0F61FC915DC0D89E5912A0024140910E0F61FC915DC0D89E5912A0024140A7052FFA0A925DC055DD239BAB0241
s:Building_1352239135666r1_0 BuildingOntology:hasRelatedSourceNumber s:RelatedSourceNumer_58772 .
s:RelatedSourceNumer_58772 BuildingOntology:wikimapiaNumber "58772" .

s:Building_1352239135666r2_0 a BuildingOntology:Building .
s:Point_1352239135666r2_1 a BuildingOntology:Point .
s:PointGeocoordinates_1352239135666r2_2 a BuildingOntology:PointGeocoordinates .
```

Model the Geospatial Data

Map the extracted geospatial data to RDF

- **Building scenario visualized on Google Earth**

Outline

- Model the Geospatial Data
- *Geospatial Data Linking*
- Geospatial Data Integration
- Conclusion and Future work

Geospatial Data Linking

Linking by the specific geospatial relationships

- **Geospatial relationships are very useful for linking**
- **Relationships often cannot be determined by just comparing the values such as name**

Distance ← executeQuery("Select ST_Distance(ST_Geography FromText(SRID=4326,s_1.getLocation),ST_GeographyFromText(SRID=4326,s_2.getLocation))")

isContained ← executeQuery("Select ST_Contains (ST_ GeographyFromText(s_1.getLocation),ST_GeographyFromText(s_2.getLocation))")

isOverlap ← executeQuery("Select ST_Overlaps(ST_Geography FromText (s_1.getLocation), ST_GeographyFromText (s_2.getLocation))")

Geospatial Data Linking

Polygon-to-Polygon:

- if (isContained=true){ similarity ← 1.0; }
- else if((isOverlap=true)) { similarity← $\left|1 - \frac{distance}{threshold}\right|$; }
- else { similarity ← ε; }

Polygon-to-Point/Point-to-Point:

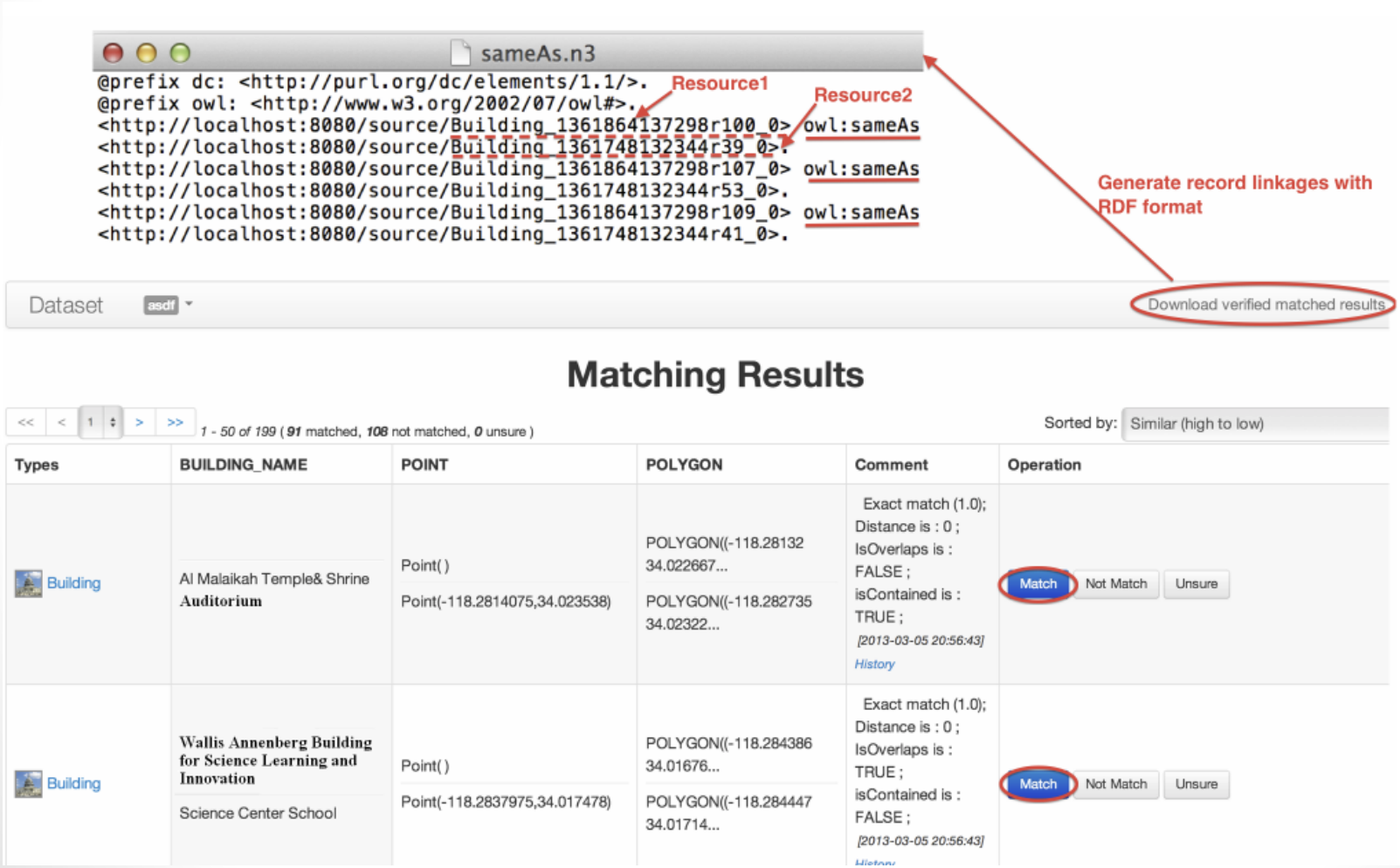
- { similarity← $\left|1 - \frac{distance}{threshold}\right|$; }

Linking Result:

- if (similarity>threshold){
- linkedPair.add(s_1,s_2);
- }

Geospatial Data Linking

Linking results illustration



Outline

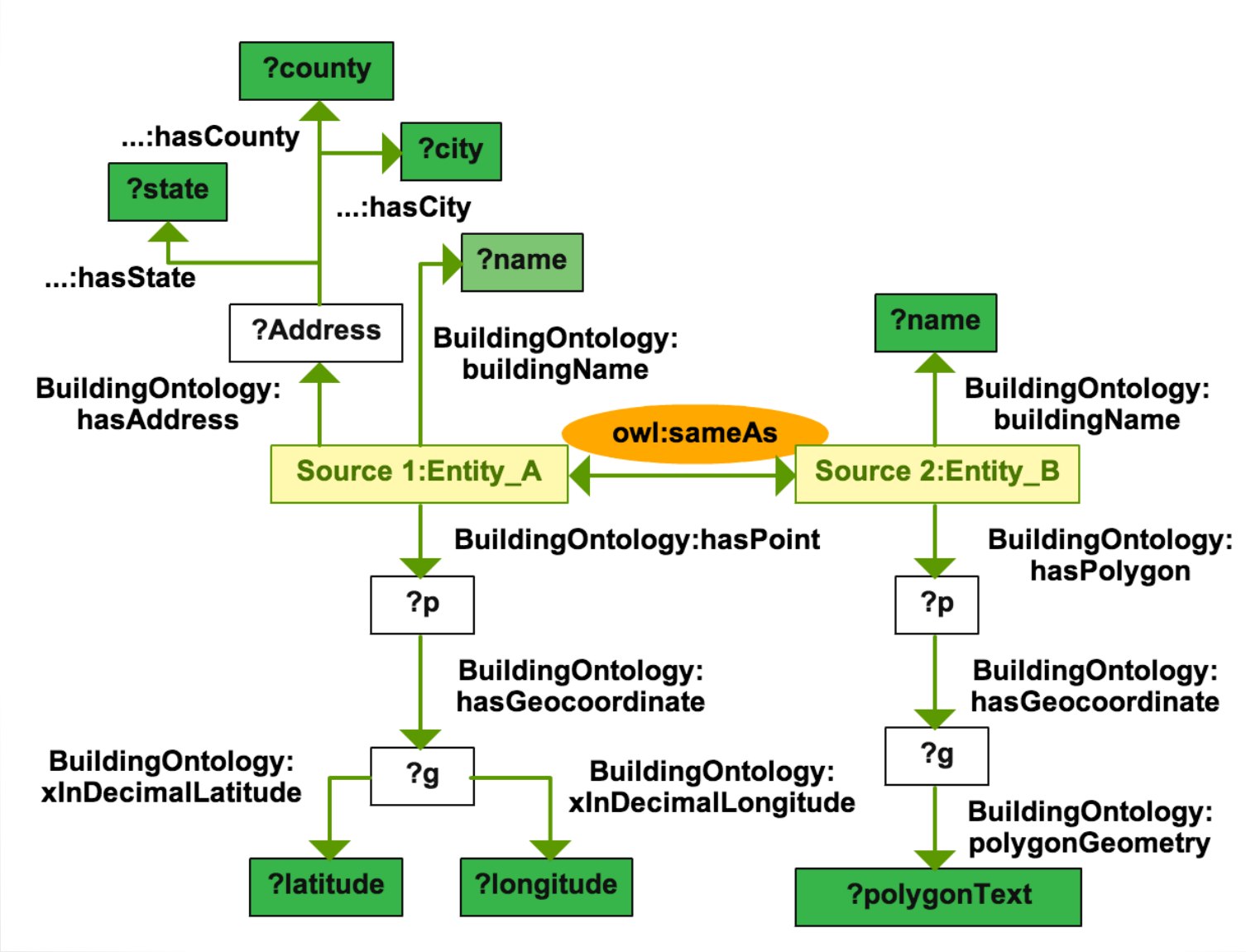
- **Model the Geospatial Data**
- **Geospatial Data Linking**
- ***Geospatial Data Integration***
- **Conclusion and Future work**

Geospatial Data Integration

Based on the record linkages, use SPARQL queries to eliminate data redundancy and combine complementary properties for integration

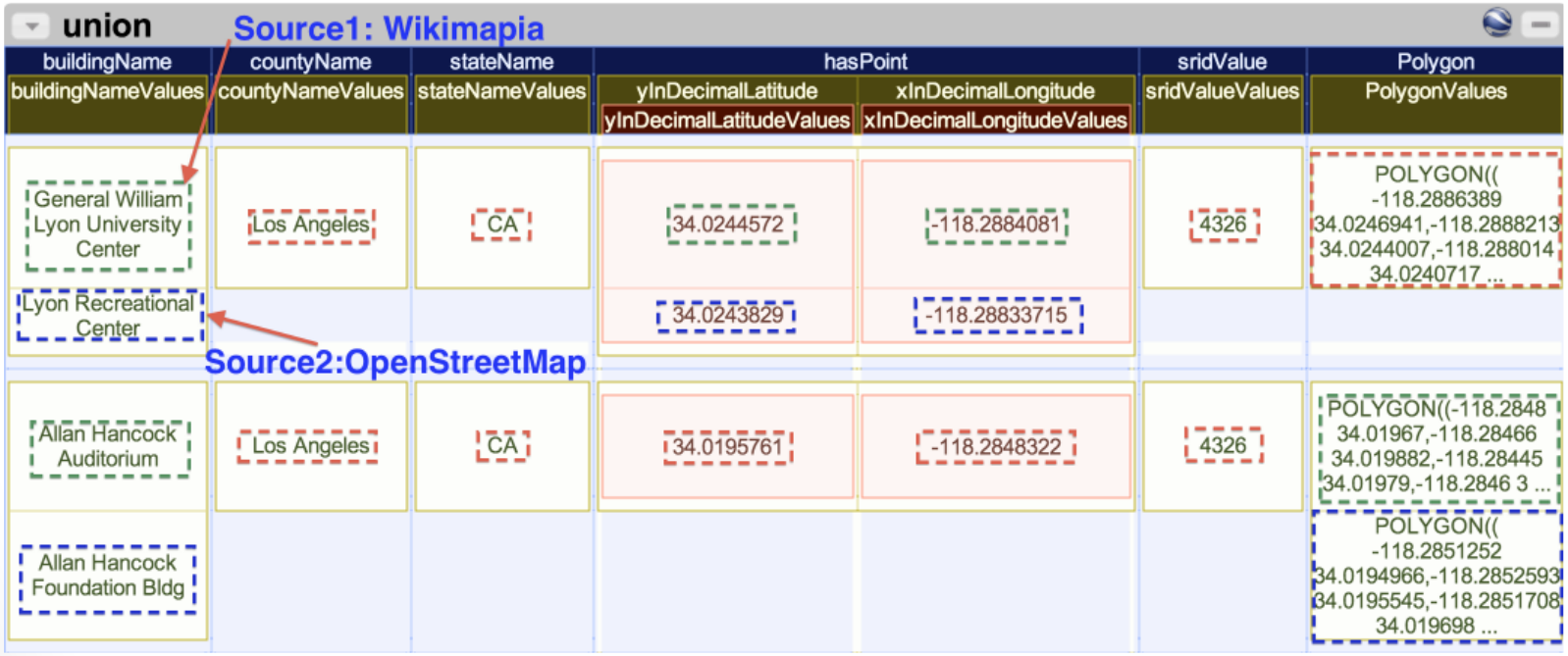
- ***sparqlS*$_1$: general query**
- **Select** discinct ?**uri**
- **Where**{
- ?uri owl:sameAs ?u.
- ?uri a BuildingOntology:Building
- }

Geospatial Data Integration



Geospatial Data Integration

Display integration results



Outline

- **Model the Geospatial Data**
- **Geospatial Data Linking**
- **Geospatial Data Integration**
- ***Conclusion and Future work***

Conclusion

We were able to empower end-users to rapidly extract, link and integrate geospatial data by means of both semantic techniques and spatial characteristics

- **Encapsulate** the retrieval algorithms as Web services
- **Align** the extracted geospatial data by mapping them to a generic geospatial ontology
- **Link** similar entities from different sources based on the matched similarity
- **Use** SPARQL queries to eliminate data redundancy and combine complementary properties

Future Work

- **Experimental comparison with other approaches**
- **Using additional attributes to optimize the geospatial data linking and integration process to improve the integration results**

Thank you!