Building Linked Data from Historical Maps

Chun Lin, Hang Su,
Craig A. Knoblock,
Yao-Yi Chiang, Weiwei Duan,
Stefan Leyk and Johannes H. Uhl

Center on Knowledge Graphs, USC/ISI
PROBLEM: RAILROADS APPEARANCE

1950

Railroads appearance in Louisville, Colorado

1965
PROBLEM: RAILROADS CHANGE

1953
Railroads change in Los Angeles, California

1966
GOALS

- Geospatial change analysis
- Utilize other knowledge sources
- Enable semantic rich queries

![Diagram showing data flow from extracted map data to geospatial linked data, connected to various knowledge sources like DBpedia and USGS.]

Center on Knowledge Graphs, USC/ISI
University of Southern California
TECHNICAL CHALLENGES

● How to determine same and different parts across maps?
● How to incrementally integrate new data?
● How to organize the data for change analysis queries?
● How to represent geospatial data in RDF?
  ○ Geometry representation
  ○ Assign unique URI
PIPELINE

Linking → Creating Linked Data → Querying
Map A and B have common and distinct segments.

Buffer out and find the common parts.

Partition line segments and mark “SameAs”.

Center on Knowledge Graphs, USC/ISI
University of Southern California
LINKING: “CONTAINS” RELATIONSHIP

3rd map comes in...

“Contains” relationship tree
LINKING: “CONTAINS” RELATIONSHIP

3rd map comes in...

“Contains” relationship tree
LINKING: RESULTS

Segmentation result
CREATING THE LINKED DATA: PREPARATION

**WKT**

<table>
<thead>
<tr>
<th>WKB</th>
<th>MULTILINESTRING((&lt;-118.157170450431 34.0000000000001,-118.15711628537 34.000117586834,-118.157055418703 34.0002624753498,-118.15701194162 34.0004160920163,......))</th>
</tr>
</thead>
<tbody>
<tr>
<td>01050000020AD1000005102000000102 000000E000000DD1FA6140F8A5D C0080000000000414048D576....</td>
<td></td>
</tr>
</tbody>
</table>

**URI**

<table>
<thead>
<tr>
<th>WKT</th>
<th>MULTILINESTRING((&lt;-118.157170450431 34.0000000000001,-118.15711628537 34.000117586834,-118.157055418703 34.0002624753498,-118.15701194162 34.0004160920163,......))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Map Source</td>
<td>USGS vector data for Los Angeles, CA</td>
</tr>
</tbody>
</table>

**Map Source**

USGS vector data for Los Angeles, CA
### RESULTS IN TABLES

<table>
<thead>
<tr>
<th>URI</th>
<th>WKT</th>
<th>Map</th>
</tr>
</thead>
<tbody>
<tr>
<td>V12-2</td>
<td>LineString…</td>
<td>A</td>
</tr>
<tr>
<td>V22-2</td>
<td>LineString…</td>
<td>B</td>
</tr>
<tr>
<td>V32</td>
<td>LineString…</td>
<td>C</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>“Contains”</th>
<th>Map vectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>URI</td>
<td>Contains</td>
</tr>
<tr>
<td>A</td>
<td>V12</td>
</tr>
<tr>
<td>B</td>
<td>V22</td>
</tr>
<tr>
<td>V12</td>
<td>V12-1</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Map metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>URI</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>“SameAs”</th>
</tr>
</thead>
<tbody>
<tr>
<td>URI</td>
</tr>
<tr>
<td>V12-2</td>
</tr>
<tr>
<td>V22-2</td>
</tr>
<tr>
<td>V32</td>
</tr>
</tbody>
</table>

Segmentation result
KARMA: A DATA INTEGRATION TOOL

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

Map vectors

“SameAs”

“Contains”

Map metadata

Center on Knowledge Graphs, USC/ISI
University of Southern California
ALTERNATIVE RAILROAD MAPS FOR LA

- USGS vector data for Los Angeles, California (373 KB)
- California Rail Network (183 KB)
- National Atlas of the United States (45 KB)
QUERYING

What railroads are in 2000 but not in 2005?

PREFIX schema: <http://schema.org/>
select distinct ?a ?mapa
where {?a schema:geo ?geo.
   ?a schema:mapType ?mapa.
   ?mapa schema:releaseDate "2000".
   filter not exists{
      ?a schema:sameAs ?b.
      ?b schema:mapType ?mapb.
      ?mapb schema:releaseDate "2005".}
minus{?a schema:contains ?x}}
RELATED WORK

- Linking process
  - Map vector data conflation [Ruiz et al. 2011]
    - Feature matching between maps
    - No segmentation on feature vectors
  - Similarity measures of vector data [Sherif et al. 2015]
    - Detailed measurement for “SameAs” with point set
    - Trade-off: computationally expensive

- Creating the Linked Data
  - Integrating geospatial information using Linked Data
    - Focuses on points of interest data, not vector
DISCUSSION

- Proposing general pipeline from linking to querying
- Integrating map vector data using Linked Data
- Making the data widely available to researchers
- Enabling the ability to answer complex queries
FUTURE WORK

- Scaling up to accommodate large volume of map data
- Optimizing Linking / Querying process
- Pipeline automation