Efficient Presentation of GeoSPARQL-Results

Bachelor‘s Thesis
Denis Veil

Examiner: Prof. Dr. Hannah Bast
Adviser: Patrick Brosi
QLever

- SPARQL engine for combined search[1]
- Knowledge base as database
  - contains statements as triples
GeoSPARQL

- Spatial extension to SPARQL for geographic information [2]
- Extending the knowledge base by spatial data [3]
- Attribute “hasGeometry” well-known text literal [4]
Well-known Text Literals

- **Point:**
  - POINT(x0 y0)

- **Line:**
  - LINESTRING(x0 y0, x1 y1, ..., xn yn)

- **Simple 2-dimensional shape:**
  - POLYGON(((x0 y0, x1 y1, ..., xn yn), (a0 b0, ...))

- **Multiple 2-dimensional shapes:**
  - MULTIPOLYGON(((x0 y0, ..., xn yn), (a0 b0, ...)), ((...)))
Well-known Text Example

- MULTIPOLYGON(((1 1, 3 1, 4 2, 2 4, 0.5 2.5), (1.5 1.5, 3 2, 2 3)), ((4 3, 4.5 4.5, 3 4)))
SPARQL Query

```sparql
PREFIX osmkey: <https://www.openstreetmap.org/wiki/Key:>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX osm: <https://www.openstreetmap.org/>

SELECT ?osm_id ?geometry WHERE {
  ?osm_id osmkey:building "university" .
  ?osm_id geo:hasGeometry ?geometry .
  FILTER contained(?envelope, "LINESTRING(7.662006 47.903578, 7.930844 48.071058)")
}
```
Motivation

- QLever delivers GeoSPARQL results
  - „all residential highways“ → 57 million objects
  - 100,000 objects ~ 35.7 MB
- Loading all objects at once
  - long request response
- Render all objects at once
  - laggy user experience
Questions?
Approach

- Create a test query from the original to get < 12,000 objects
- Cluster the result
  - optional: create sub clusters
- Render the result with Leaflet[^5]
Flow Chart 2/2

1. **Load URL**
   - `load(url)`
   - **Result**
   - **Parse result to objects**
   - **Calculate (sub-)clusters**
   - **Show clusters**
   - If loaded objects < result size (complete)

2. **Zoom in and load new area**
   - **map.on(double-click)**
   - **If ≤ MAX_OBJECTS**
   - **Calculate envelope bounds**
   - **checkBounds()**
   - **map.on(drag-end)**
   - If loaded objects < result size (complete)
Clustering

- K-means clustering
  - \( k = \lfloor \log_2(\text{length of array}) \rfloor \)

- Centroid initialization
  - custom “Naïve Sharding”[6]

- Create sub-clusters if a cluster has too many objects
Clustering Example 1/4
Clustering Example 2/4
Clustering Example 3/4
Clustering Example 4/4
Loading New Area 1/3

- northwest
- north
- northeast

- west
- visible map on screen
- east

- southwest
- south
- southeast
<table>
<thead>
<tr>
<th>northwest</th>
<th>north</th>
<th>northeast</th>
<th>new area</th>
</tr>
</thead>
<tbody>
<tr>
<td>west</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>southwest</td>
<td>south</td>
<td>southeast</td>
<td>new area</td>
</tr>
</tbody>
</table>

Visible map on screen
<table>
<thead>
<tr>
<th>new area</th>
<th>new area</th>
<th>new area</th>
<th>new area</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

visible map on screen
Questions?
## Evaluation

<table>
<thead>
<tr>
<th></th>
<th>memory (MB)</th>
<th>response time (s) (50Mbit DL)</th>
<th>render time (s)</th>
<th>time total (s)</th>
<th>#objects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MarkerCluster</td>
<td>19,7</td>
<td>2,9</td>
<td>7,6</td>
<td>10,5</td>
<td>50.000</td>
</tr>
<tr>
<td>Cluster only</td>
<td></td>
<td></td>
<td>3,6</td>
<td>6,5</td>
<td></td>
</tr>
<tr>
<td><strong>Baseline</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MarkerCluster</td>
<td>35,7</td>
<td>5,4</td>
<td>26,2</td>
<td>29,1</td>
<td>100.000</td>
</tr>
<tr>
<td>Cluster only</td>
<td></td>
<td></td>
<td>7</td>
<td>9,9</td>
<td></td>
</tr>
<tr>
<td>Filter + Cluster</td>
<td>0,3 – 2,4</td>
<td>0,2 - 0,8</td>
<td>0,2 - 0,6</td>
<td>0,4 - 1,4</td>
<td>950 - 7983</td>
</tr>
</tbody>
</table>
Conclusion

- Average loading time is < 1sec
- Average rendering time is < 1sec
- Loading new area almost instantly
Future Work

- Extending spatial data on QLever
  - center of the object
  - tile and zoom level of the center
- Pre-compute clusters on QLever
  - send only clusters on lower zoom levels
  - on click send subclusters and/or objects
- Improve the cluster shape
Thank you for your attention
Sources


[2] GeoSPARQL - A Geographic Query Language for RDF Data

Sources


## Appendix: Evaluation

<table>
<thead>
<tr>
<th></th>
<th>response time (ms)</th>
<th>render time (ms)</th>
<th>memory (MB)</th>
<th>#objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (simple loading)</td>
<td>1461.4</td>
<td>2132</td>
<td>9.5 MB</td>
<td>25000</td>
</tr>
<tr>
<td></td>
<td>2920.0</td>
<td>4727.9</td>
<td>19.7 MB</td>
<td>50000</td>
</tr>
<tr>
<td></td>
<td>4121.2</td>
<td>6426.6</td>
<td>28.4 MB</td>
<td>75000</td>
</tr>
<tr>
<td></td>
<td>5401.8</td>
<td>8107.3</td>
<td>35.7 MB</td>
<td>100000</td>
</tr>
<tr>
<td>Leaflet.MarkerCluster</td>
<td>1432.7</td>
<td>2798.7</td>
<td>9.5 MB</td>
<td>25000</td>
</tr>
<tr>
<td></td>
<td>2872.5</td>
<td>7654.0</td>
<td>19.7 MB</td>
<td>50000</td>
</tr>
<tr>
<td></td>
<td>4134.2</td>
<td>14511.3</td>
<td>28.4 MB</td>
<td>75000</td>
</tr>
<tr>
<td></td>
<td>5465.2</td>
<td>26200.3</td>
<td>35.7 MB</td>
<td>100000</td>
</tr>
<tr>
<td>Cluster only</td>
<td>1480.3</td>
<td>1816.5</td>
<td>9.5 MB</td>
<td>25000</td>
</tr>
<tr>
<td>(own implementation)</td>
<td>2884.1</td>
<td>3600.2</td>
<td>19.7 MB</td>
<td>50000</td>
</tr>
<tr>
<td></td>
<td>4103.5</td>
<td>5170.1</td>
<td>28.4 MB</td>
<td>75000</td>
</tr>
<tr>
<td></td>
<td>5417.5</td>
<td>6971.3</td>
<td>35.7 MB</td>
<td>100000</td>
</tr>
<tr>
<td>Filter + Cluster</td>
<td>196.8</td>
<td>206.1</td>
<td>302 kB</td>
<td>950</td>
</tr>
<tr>
<td>(own implementation)</td>
<td>289.1</td>
<td>293.0</td>
<td>861 kB</td>
<td>2786</td>
</tr>
<tr>
<td></td>
<td>743.6</td>
<td>601.3</td>
<td>2.4 MB</td>
<td>7983</td>
</tr>
<tr>
<td></td>
<td>692.4</td>
<td>414.7</td>
<td>2.4 MB</td>
<td>7983</td>
</tr>
<tr>
<td></td>
<td>706.0</td>
<td>589.9</td>
<td>2.4 MB</td>
<td>7983</td>
</tr>
</tbody>
</table>