RDF Data Analysis with Activation Patterns
Graz University of Technology
IAIK
Peter Teufl/Günther Lackner
Overview

- RDF and queries
  - RDF
  - CIA Worldfactbook as RDF data
  - Query Languages SPARQL/SERQL
  - Fuzzy Queries

- Activation Patterns
  - Activation Patterns
  - Techniques
  - Generation
  - Applications

- Example
  - CIA Worldfactbook as RDF data
  - AP Generation
  - Semantic Relations
  - Semantic Search
  - Clustering
CIA World Factbook

**Location:**
Central Europe, north of Italy and Slovenia

**Geographic coordinates:** 47°20'N, 13°20'E

**Land boundaries:**
Total: 2,362 km, border countries: Czech Republic 362 km, Germany 704 km, Hungary 365 km, Italy 430 km, Liechtenstein 35 km, Slovakia 91 km, Slovenia 330 km, Switzerland 164 km

**Area:**
- Total: 83,871 sq km
- Country comparison to the world: 111
- Land: 82,445 sq km
- Water: 1,426 sq km

**External International Agreements:**
- signed, but not ratified: none of the selected agreements
RDF Data Extraction

- SQL like query languages
- Extracting information about triples
- SPARQL, SERQL
Fuzzy queries?

- How is literacy rate typically related to the unemployment rate?
- Retrieve all countries according to their similarity to Austria
- Find the typical features for countries that export bananas and retrieve all countries that have similar features but do not export bananas themselves
- Cluster countries according to features related to gross domestic product and imported commodities
Activation Patterns

- Machine Learning
- Artificial intelligence
- One solution for everything
- Clustering, semantic aware search queries, supervised learning, anomaly detection, knowledge extraction
- Combine semantic networks, spreading activation and machine learning
Activation Pattern Techniques

AI and Machine Learning

- Spreading Activation
- Unsupervised Learning
  - Neural Gas Family
  - SOM Family
  - HAC
- Supervised Learning
  - Decision Trees
  - Neural Networks
  - Support Vector Machines

- Similarity Measures
  - Euclidean
  - Cosine
- Associative Networks
- Semantic Networks
Activation Patterns

- L1: Feature Extraction from RDF
- L2: Node generation
- L3: Network generation
- L4: SA
- L5: Analysis
Activation Patterns: L5

- Supervised learning
- Unsupervised learning
- Anomaly detection
- Feature relevance
- Relations
- Semantic search
- Activation energy
- Interconnections
- Connections between nodes
- Search for related Patterns
- Create fingerprints
- Categorize data
- Activation Patterns
- Techniques
- Generation
- Applications

Search query
Instance from data set
All instances (clustering)
Single feature
New instance
Activate nodes
Spread activation

Search query
Instance from data set
All instances (clustering)
Single feature
New instance
Activate nodes
Spread activation
Applications

- Malware Analysis
- Event Correlation
- RDF Analysis
- Twitter Mining
- User Tracking in WIFI Networks
- e-Participation
  - Text analysis
Dataset

- CIA World Factbook
- Common knowledge
- Symbolic/real features
- Analysis procedures easy to verify
- Well known features: literacy rate, exports, unemployment etc.
Analysis - Steps

- L1: Extract country features
- L2/L3: Create nodes/links
- L4: SA
- L5: Analysis
Semantic Relations

- `mapReference: Africa`
- `mapReference: Europe`
- `exports: crude oil`
- `exports: machinery, chemicals`

### Relation 1

<table>
<thead>
<tr>
<th></th>
<th><code>mapReference: Africa</code></th>
<th><code>mapReference: Europe</code></th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment (%)</td>
<td>24.45 (1.0)</td>
<td>04.02 (1.0)</td>
</tr>
<tr>
<td>Literacy Total (%)</td>
<td>41.75 (1.0)</td>
<td>95.92 (1.0)</td>
</tr>
<tr>
<td>Gross Agriculture (%)</td>
<td>40.41 (1.0)</td>
<td>03.68 (1.0)</td>
</tr>
<tr>
<td>Gross Services (%)</td>
<td>37.45 (1.0)</td>
<td>70.37 (1.0)</td>
</tr>
<tr>
<td>Gross Industry (%)</td>
<td>21.24 (1.0)</td>
<td>21.24 (1.0)</td>
</tr>
<tr>
<td>Exports</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coffee (%)</td>
<td>52.87 (0.4)</td>
<td>60.15 (0.8)</td>
</tr>
<tr>
<td>Cotton (%)</td>
<td>14.80 (0.8)</td>
<td>14.80 (0.2)</td>
</tr>
<tr>
<td>Chemicals (%)</td>
<td>30.39 (1.0)</td>
<td>21.24 (0.4)</td>
</tr>
</tbody>
</table>

### Relation 2

<table>
<thead>
<tr>
<th></th>
<th><code>exports: crude oil</code></th>
<th><code>exports: machinery, equipment</code></th>
<th><code>exports: chemicals</code></th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment (%)</td>
<td>24.45 (1.0)</td>
<td>04.02 (1.0)</td>
<td>12.93 (0.7)</td>
</tr>
<tr>
<td>Literacy Total (%)</td>
<td>80.06 (1.0)</td>
<td>95.92 (1.0)</td>
<td>80.06 (1.0)</td>
</tr>
<tr>
<td>Gross Agriculture (%)</td>
<td>03.68 (1.0)</td>
<td>03.68 (1.0)</td>
<td>14.80 (0.3)</td>
</tr>
<tr>
<td>Gross Services (%)</td>
<td>48.77 (1.0)</td>
<td>70.37 (1.0)</td>
<td>60.15 (0.8)</td>
</tr>
<tr>
<td>Gross Industry (%)</td>
<td>41.40 (1.0)</td>
<td>30.39 (1.0)</td>
<td>41.40 (0.3)</td>
</tr>
<tr>
<td>Exports</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crude Oil (%)</td>
<td>52.87 (0.4)</td>
<td>30.39 (1.0)</td>
<td>0.1</td>
</tr>
<tr>
<td>Coffee (%)</td>
<td>14.80 (0.8)</td>
<td>14.80 (0.2)</td>
<td>1.0</td>
</tr>
<tr>
<td>Machinery and Equipment</td>
<td>30.39 (1.0)</td>
<td>30.39 (1.0)</td>
<td></td>
</tr>
<tr>
<td>Chemicals (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Semantic Search

- Activate node “crude oil”
- Search for similar countries
- Best matching: countries that export crude oil
- BUT: countries that do not export crude oil, but are similar, are also retrieved

<table>
<thead>
<tr>
<th>Result</th>
<th>Country</th>
<th>exports</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Equatorial Guinea</td>
<td>timber, cocoa, petroleum</td>
</tr>
<tr>
<td>13</td>
<td>Congo</td>
<td>lumber, cocoa, petroleum</td>
</tr>
<tr>
<td>14</td>
<td>Kuwait</td>
<td>fertilizers, oil and refined products</td>
</tr>
<tr>
<td>15</td>
<td>Cameroon</td>
<td>lumber, cotton, petroleum products</td>
</tr>
<tr>
<td>16</td>
<td>Qatar</td>
<td>petroleum products, fertilizers, steel</td>
</tr>
<tr>
<td>202</td>
<td>Germany</td>
<td>chemicals, textiles, foodstuffs</td>
</tr>
</tbody>
</table>
Cluster countries according to the distribution of the gross domestic product (agriculture, industry, services)
Conclusions

- Anomaly detection, feature relevance, supervised learning
- Activation Patterns: Basis for arbitrary analysis
- Enable fuzzy querying within the semantic net
- JAVA Framework under development
- Application domain: arbitrary data
Thank You!

Looking forward to answering your questions!

- [http://www.iaik.tugraz.at](http://www.iaik.tugraz.at)
- [peter.teufl@iaik.tugraz.at](mailto:peter.teufl@iaik.tugraz.at)
- [guenther.lackner@studio78.at](mailto:guenther.lackner@studio78.at)