SIMILARITY-BASED INFORMATION RETRIEVAL AND ITS ROLE WITHIN SPATIAL DATA INFRASTRUCTURES

Krzysztof Janowicz\textsuperscript{1}, Marc Wilkes\textsuperscript{1}, and Michael Lutz\textsuperscript{2}

\textsuperscript{1} Institute for Geoinformatics, University of Muenster, Germany
{janowicz\textbar marc.wilkes}@uni-muenster.de

\textsuperscript{2} European Commission – Joint Research Centre
Institute for Environment and Sustainability, Ispra, Italy
michael.lutz@jrc.it

GIScience 2008
1 A Walk in the Park

2 Semantics-based Information Retrieval

3 Information Retrieval Paradigms

4 User Interfaces

5 ... Leaving the Park
Sometime it Needs a Walk in the Park
The ‘Similarity-based Retrieval’ - Park
1. A Walk in the Park

2. Semantics-based Information Retrieval

3. Information Retrieval Paradigms

4. User Interfaces

5. ... Leaving the Park
Some Key Terms

- The **intended concept** \((C_i)\) represents exactly the information the user is looking for (without a direct mapping in the queried ontology).

- The **search concept** \((C_s)\) is the concept actually used in the search (intensional search).

- The **target concepts** \((C_t)\) are the set of concepts to which the search is applied.

- The **reference individuals** \((I_r)\) allow for a *query-by-example* (extensional search).

→ The result of a query, i.e., the concepts or individuals returned to the user, is the better the more accurate \(C_s\) approximates \(C_i\). In the ideal case, \(C_s\) would be equal to \(C_i\).
Subsumption and Similarity-based Retrieval
**Subsumption and Similarity-based Retrieval**

- **WaterBody** \(\equiv \ldots \cap \text{HydrographicFeature} \cap \ldots\)
- **Canal** \(\equiv \ldots \cap \text{WaterBody} \cap (\geq 2 \text{ connects.Waterbody}) \cap \ldots\)
- **River** \(\equiv \ldots \cap \text{WaterBody} \cap (\exists \text{hasOrigin.Spring}) \cap (\exists \text{hasDestination.Waterbody}) \cap \ldots\)
**Subsumption and Similarity-based Retrieval**

- **Canal** ≡ \ldots \, \cap \, WaterBody \, \cap \, (\geq 2 \text{ connects. Waterbody}) \, \cap \, \ldots
- **River** ≡ \ldots \, \cap \, WaterBody \, \cap \, (\exists \text{hasOrigin. Spring}) \, \cap \, (\exists \text{hasDestination. Waterbody}) \, \cap \, \ldots
- **ABC** ≡ \ldots \, \cap \, XYZ \, \cap \, (\geq 2 \text{ connects. Waterbody}) \, \cap \, \ldots
SUBSUMPTION AND SIMILARITY-BASED RETRIEVAL

'Compare different types of water bodies for similarity.'

How to integrate the context concept into a user interface?
1 A Walk in the Park

2 Semantics-based Information Retrieval

3 Information Retrieval Paradigms

4 User Interfaces

5 ... Leaving the Park
SUBSUMPTION-BASED INFORMATION RETRIEVAL PARADIGMS

a) \( C_s \subseteq C_i \)

b) \( C_i \subseteq C_s \)

c) \( C_s \equiv C_1 \cap C_2 \cap C_3 \)

d) \( C_s \equiv C_1 \cup C_2 \cup C_3 \cup C_4 \)
SIMILARITY-BASED INFORMATION RETRIEVAL PARADIGMS

e) 

f) 

SIMILARITY-BASED INFORMATION RETRIEVAL PARADIGMS

g) 

h)
A Walk in the Park

Semantics-based Information Retrieval

Information Retrieval Paradigms

User Interfaces

... Leaving the Park
INTER-CONCEPT SIMILARITY-BASED RETRIEVAL

You are currently looking for features with

any words in the name, as feature type, and that are located within this region:

<table>
<thead>
<tr>
<th>feature type suggestions</th>
<th>supertype(s)</th>
<th>similar types</th>
</tr>
</thead>
<tbody>
<tr>
<td>River</td>
<td>Watercourse</td>
<td>Canal, Irrigation, Canal, Lake, Reservoir, Tarn, Inlet, Ocean, Groundwater, Sewage</td>
</tr>
</tbody>
</table>

14 features found

- Lake Fork River - Utah, US
- Right Hand Fork East Bear River - Utah, US
- Little South Fork Provo River - Utah, US
- Duchesne River - Utah, US
- East Fork Bear River - Utah, US
- Left Hand Fork East Fork Bear River - Utah, US
- Provo River - Utah, US
- Weber River - Utah, US
- East Fork Duchesne River - Utah, US
- West Fork Duchesne River - Utah, US

demo & source code: http://sim-dl.sourceforge.net/applications/
Inter-Instance Similarity-based Retrieval

demo & source code: http://sim-dl.sourceforge.net/applications/
COMBINED SIMILARITY-BASED RETRIEVAL

Selected reference features:
- Esnagami River - Ontario, CDN
- Rideau Canal - Ontario, CDN
- Marmion Lake - Ontario, CDN

Search within this region:

1. Add reference feature
2. Selected reference features
3. Search bar

14 features found:
- Lake Fork River - Utah, US
- Right Hand Fork East Bear River - Utah, US
- Little South Fork Provo River - Utah, US
- Duchesne River - Utah, US
- East Fork Bear River - Utah, US
- Left Hand Fork East Fork Bear River - Utah, US
- Provo River - Utah, US
- Weber River - Utah, US
- East Fork Duchesne River - Utah, US
- West Fork Duchesne River - Utah, US

Not implemented yet.
1 A Walk in the Park

2 Semantics-based Information Retrieval

3 Information Retrieval Paradigms

4 User Interfaces

5 ... Leaving the Park
Spatial Data Infrastructures?

Similarity-based Information Retrieval

and its Role within Spatial Data Infrastructures
Spatial Data Infrastructures

... integration into the Geospatial Semantic Web.
QUESTIONS?

Thank you!

SimCat Project Website: http://sim-dl.sourceforge.net

The SIM-DL server, plug-in and user interfaces are free and open source software and can be downloaded at the SimCat project website.