$SIM - DL_A$: A Novel Semantic Similarity Measure for Description Logics Reducing Inter-Concept to Inter-Instance Similarity

Krzysztof Janowicz and Marc Wilkes

Münster Semantic Interoperability Lab (MUSIL)
Institute for Geoinformatics
University of Muenster, Germany
{janowicz|marc.wilkes}@uni-muenster.de

June 4, 2009

1 SEMANTICS IN GISCIENCE
- Geoinformation
- Spatial Data Infrastructure
- Semantic Enablement Layer

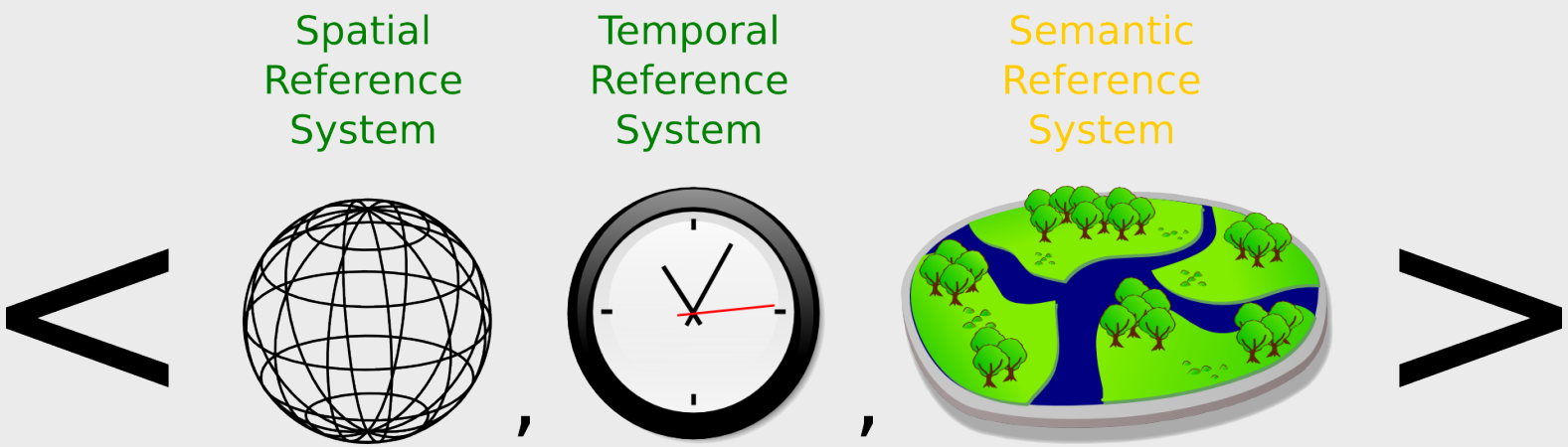
2 SIMILARITY (REASONING) IN GISCIENCE
- ADL Gazetteer Interface
- SimCat Gazetteer Interface
- SIM-DL Server and Plugin

3 $SIM - DL_A$
- Why $SIM - DL_A$?
- The Similarity Tableau for $\mathcal{SHI}$
- A Simplified Example

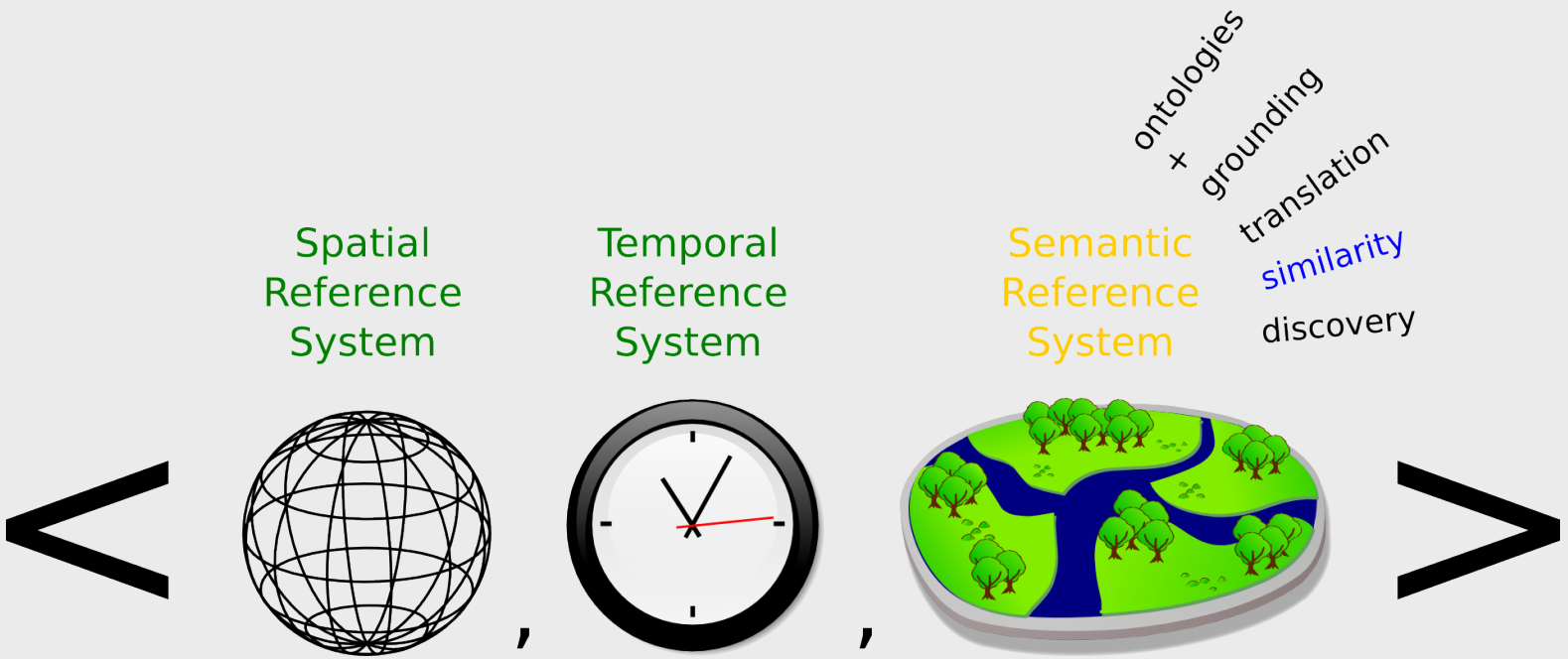
4 DISCUSSION & OUTLOOK

GEOINFORMATION

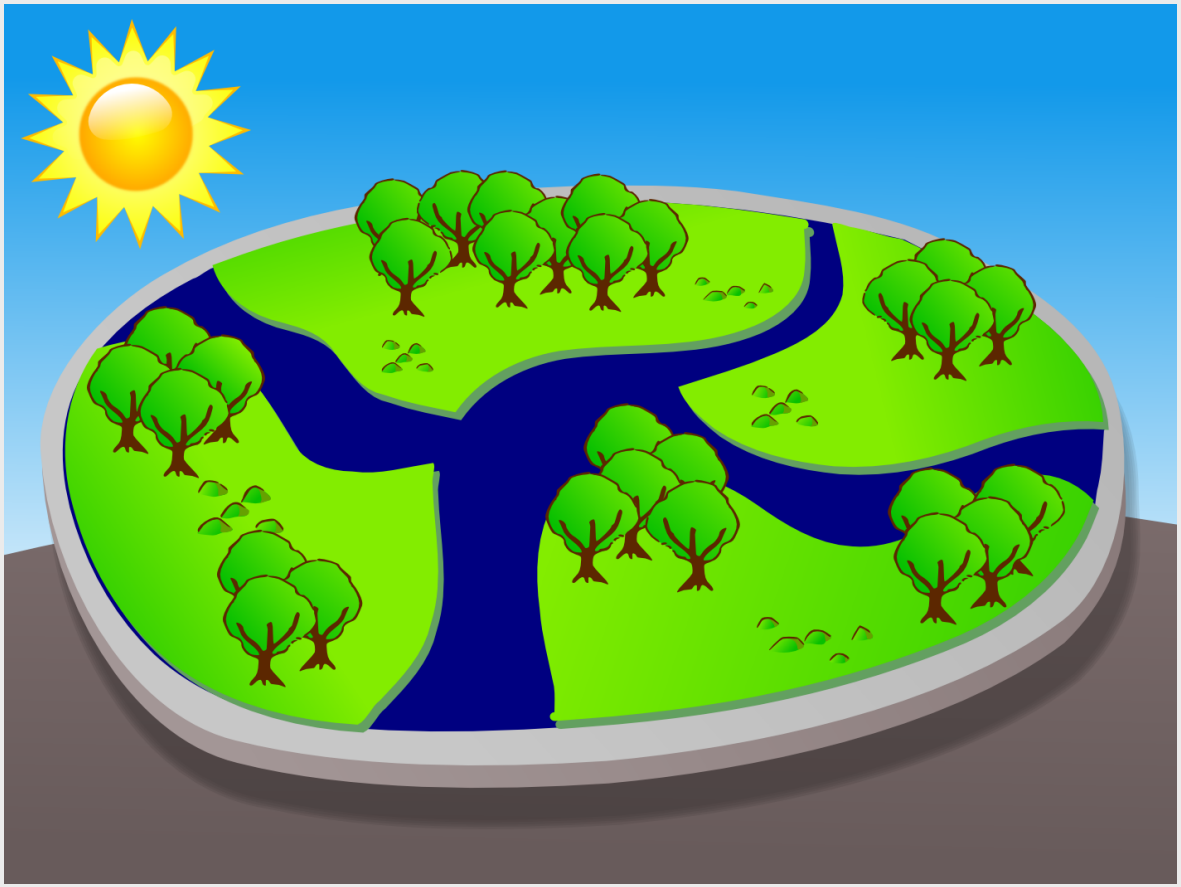
Geoinformation



GEOINFORMATION

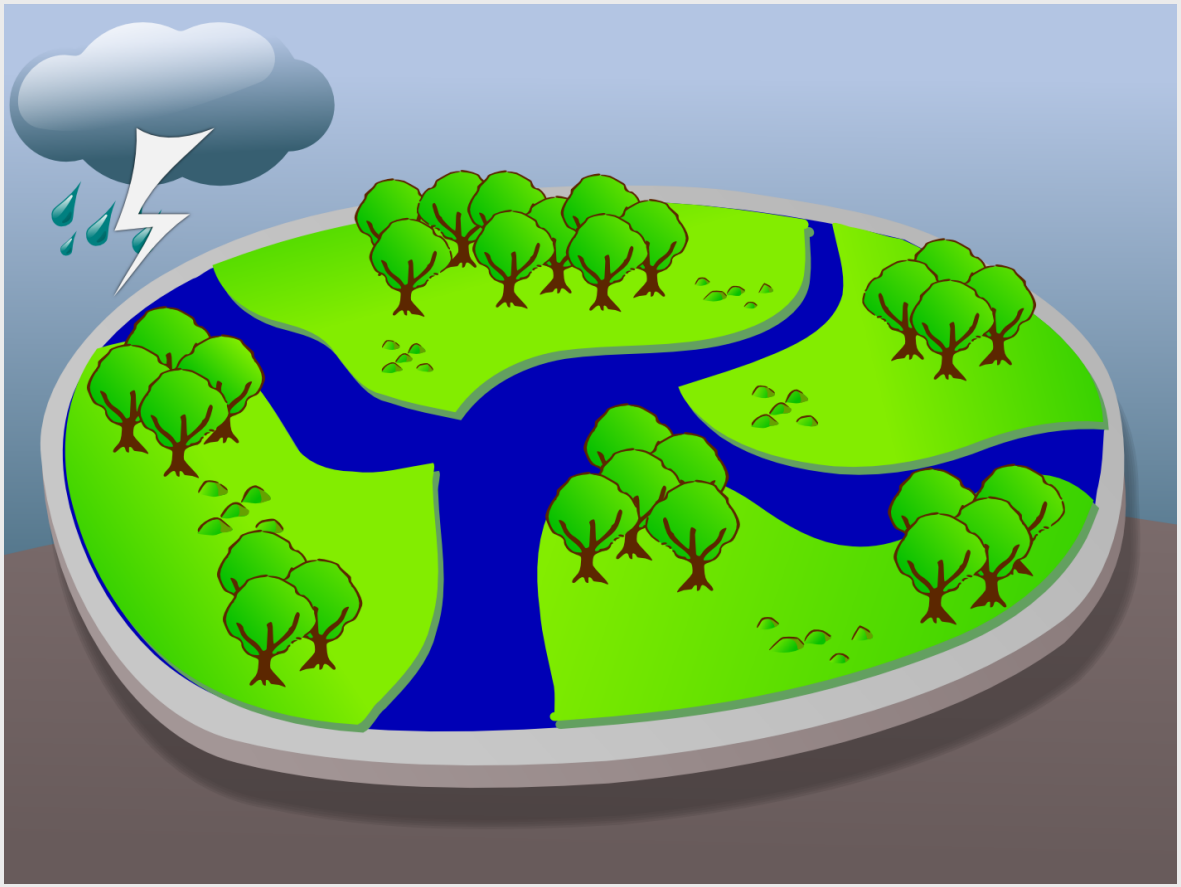


SPATIAL DATA INFRASTRUCTURE: OGC SERVICES



- Show me Hills, Forests, Rivers, *[Feature Type]* ... at *[Loc]* → WFS
- Show me a map of this area → WMS

SPATIAL DATA INFRASTRUCTURE: OGC SERVICES

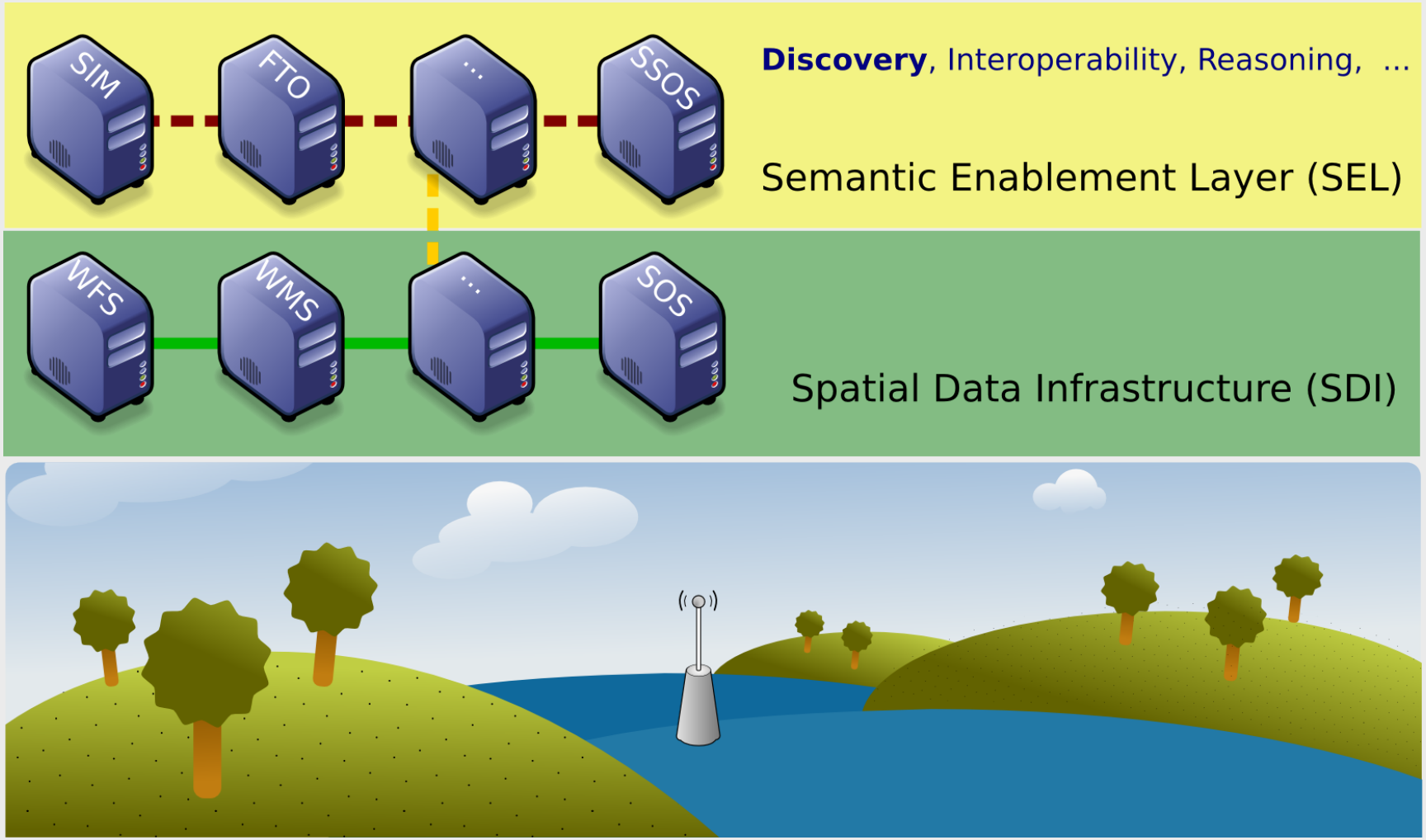


- Water gauge? *[Stimulus Type]* ... at *[Loc][Time]* → SOS → WPS
- *[Sensor]* move to *[Loc][Time]* → SPS

SPATIAL DATA INFRASTRUCTURE: OGC SERVICES

- Flooding! *[Event Type]* ... at *[Loc][Time]* → SAS → WNS

YET ANOTHER LAYER CAKE



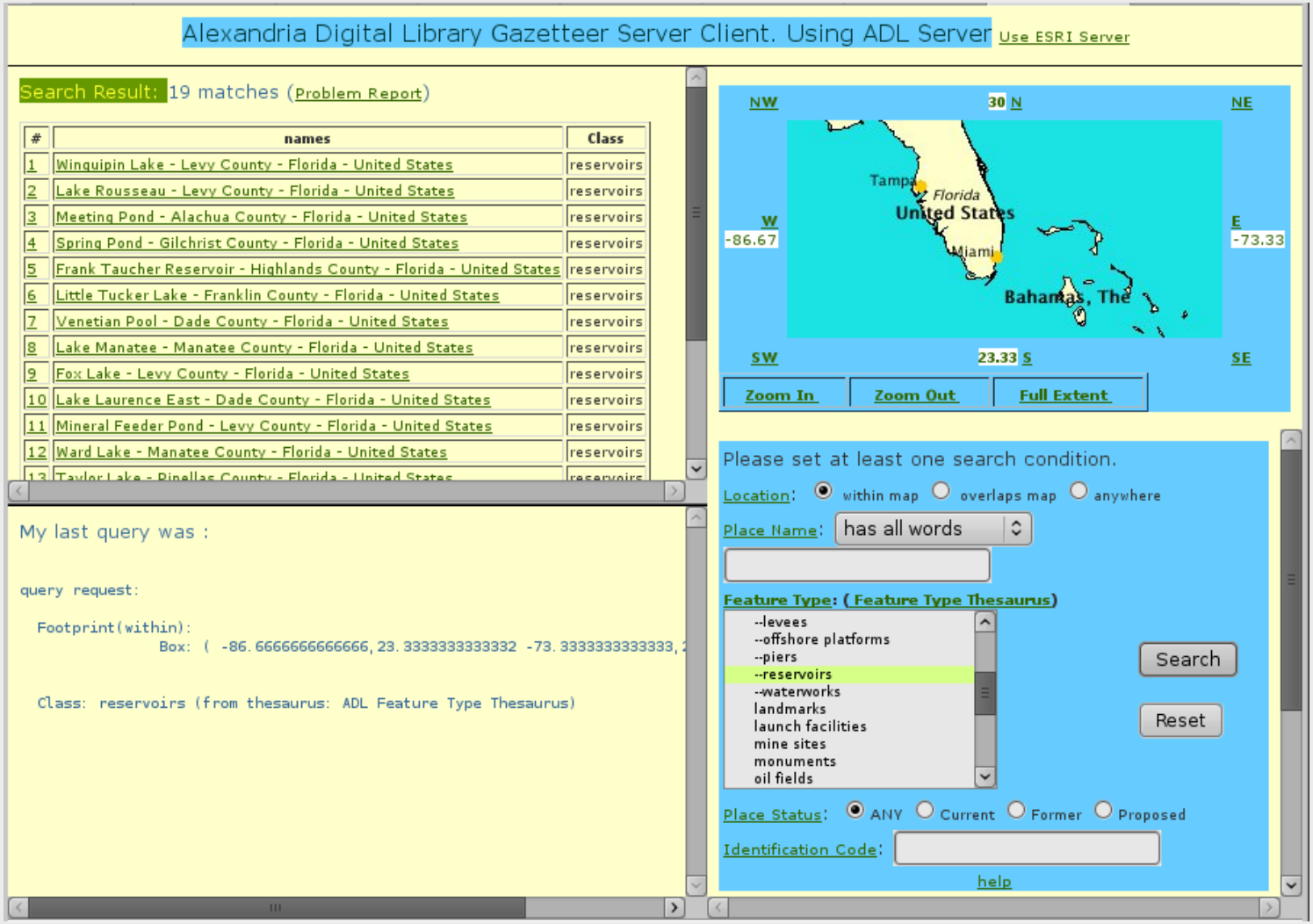
1 Semantics in GIScience
- Geoinformation
- Spatial Data Infrastructure
- Semantic Enablement Layer

2 Similarity (Reasoning) in GIScience
- ADL Gazetteer Interface
- SimCat Gazetteer Interface
- SIM-DL Server and Plugin

3 $SIM - DL_A$
- Why $SIM - DL_A$?
- The Similarity Tableau for $\mathcal{SHI}$
- A Simplified Example

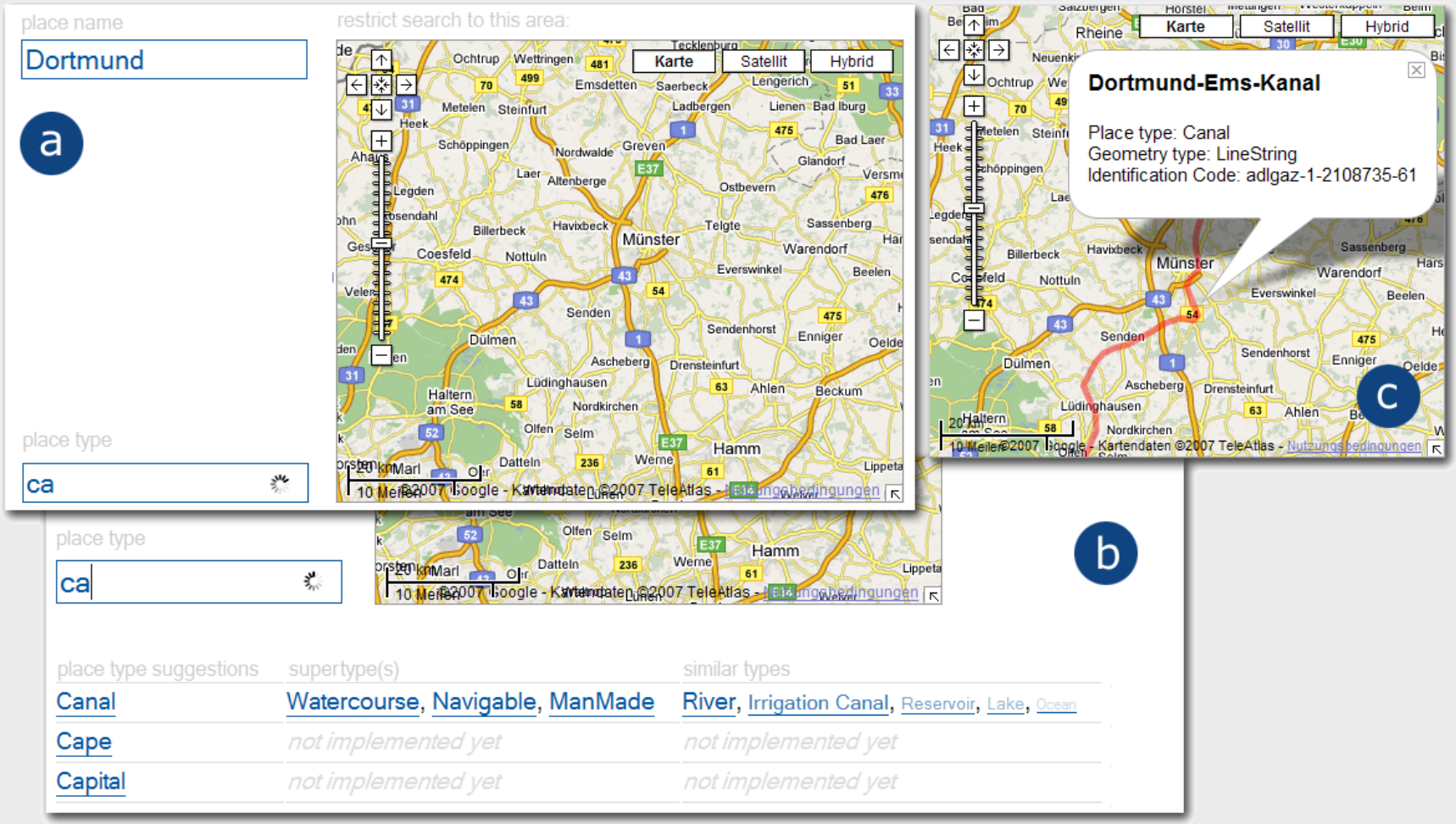
4 Discussion & Outlook

ADL GAZETTEER INTERFACE



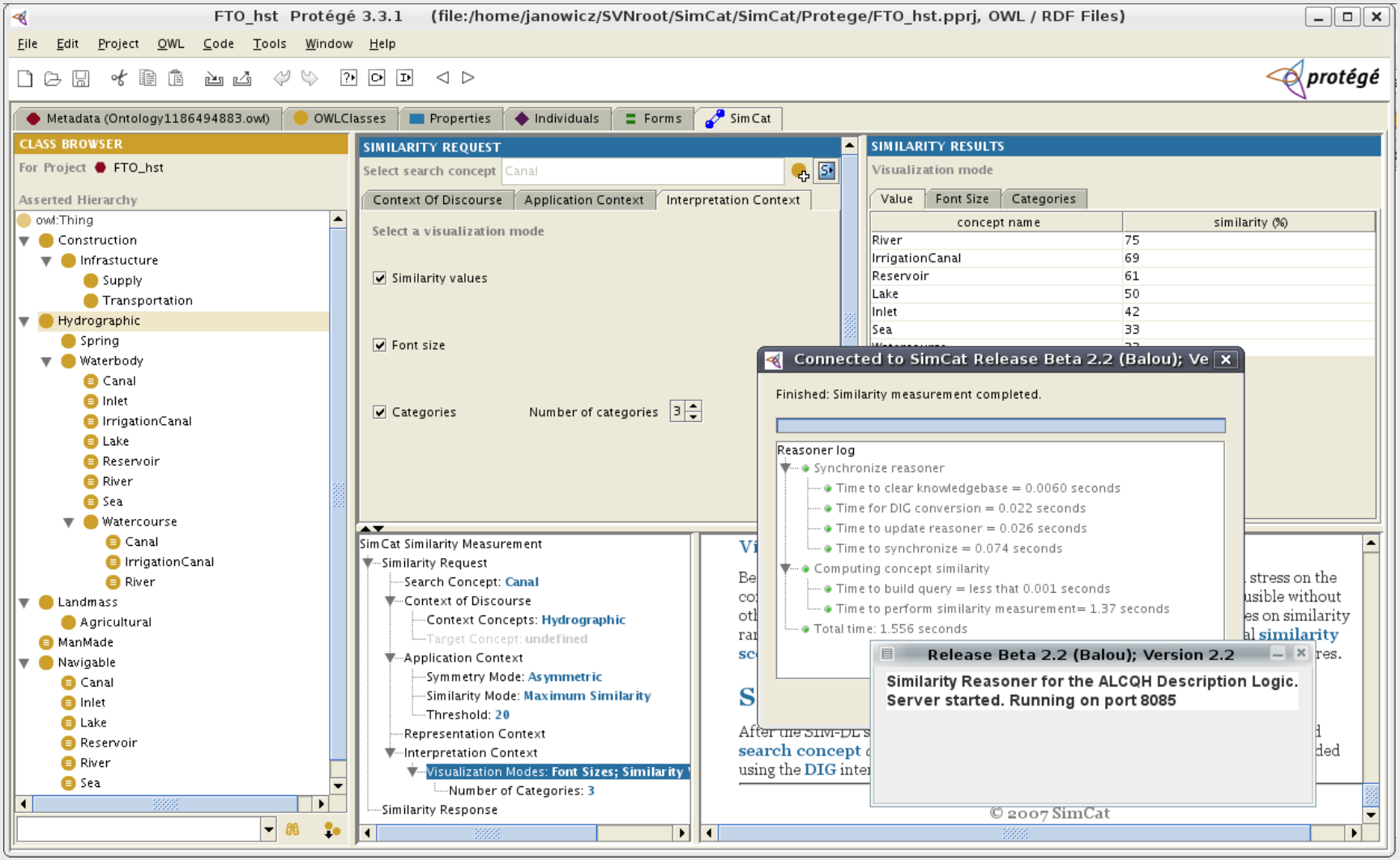
- Based on a Feature Type Thesaurus (> 1000 types)

SimCat Gazetteer Interface



- Based on a Feature Type Ontology

SIM-DL SERVER AND PLUGIN



1 SEMANTICS IN GISCIENCE
- Geoinformation
- Spatial Data Infrastructure
- Semantic Enablement Layer

2 SIMILARITY (REASONING) IN GISCIENCE
- ADL Gazetteer Interface
- SimCat Gazetteer Interface
- SIM-DL Server and Plugin

3 $SIM - DL_A$
- Why $SIM - DL_A$?
- The Similarity Tableau for $\mathcal{SHI}$
- A Simplified Example

4 DISCUSSION & OUTLOOK

The Challenge

- SIM-DL is a structural measure; it requires the normalization of concept descriptions and recursively compares the (syntactic) structure of the normalized descriptions for similarity.
- Feature type ontologies require expressive description logics such as OWL-DL; a structural approach fails to support these DLs.

- Why not use an ABox-based similarity measure?
- Most feature type ontologies cannot be populated so far. Features are stored in shape-files, geo-databases, GML files, satellite images,...

$\longrightarrow$ **Can we learn from the history of subsumption reasoning and switch to a tableau-based approach?**

The Idea

⟶ Reduce the problem of inter-concept similarity to inter-instance similarity?

- How to get or compute the instances required for comparison?
- From the completion tree of a modified tableau algorithm as used for satisfiability checking. Instead of trying to find one (clash-free) model, generate a set of proxy individuals for comparison.

- How to compute similarity between proxy individuals?
- Reuse existing similarity framework, i.e., functions, contexts, similarity modes, and the alignment procedure from SIM-DL.
  - Jaccard similarity coefficient of co-occurrence
  - Network measures for role hierarchies and neighborhood
  - ...

The Similarity Tableau for $\mathcal{SHI}$

...

The $\sqcup$-rule:

Condition: $C_1 \sqcup C_2 \in \mathcal{L}(x)$ and x is not indirectly blocked.

Action: Create three $\sqcup$-successors w, y, z of x with:

$\mathcal{L}(w) := (\mathcal{L}(x) \backslash \{C_1 \sqcup C_2\}) \cup \{C_1\}$

$\mathcal{L}(y) := (\mathcal{L}(x) \backslash \{C_1 \sqcup C_2\}) \cup \{C_2\}$

$\mathcal{L}(z) := (\mathcal{L}(x) \backslash \{C_1 \sqcup C_2\}) \cup \{C_1, C_2\}$

...

The $\forall$-rule:

Condition: $(\forall R.C) \in \mathcal{L}(x)$, x is not indirectly blocked.

Action:

If there is an R-neighbor y of x and $C \notin \mathcal{L}(y)$:

$\mathcal{L}(y) := \mathcal{L}(y) \cup \{C\}$.

If there is no R-neighbor y of x, create two $\forall$-successors y, z of x with:

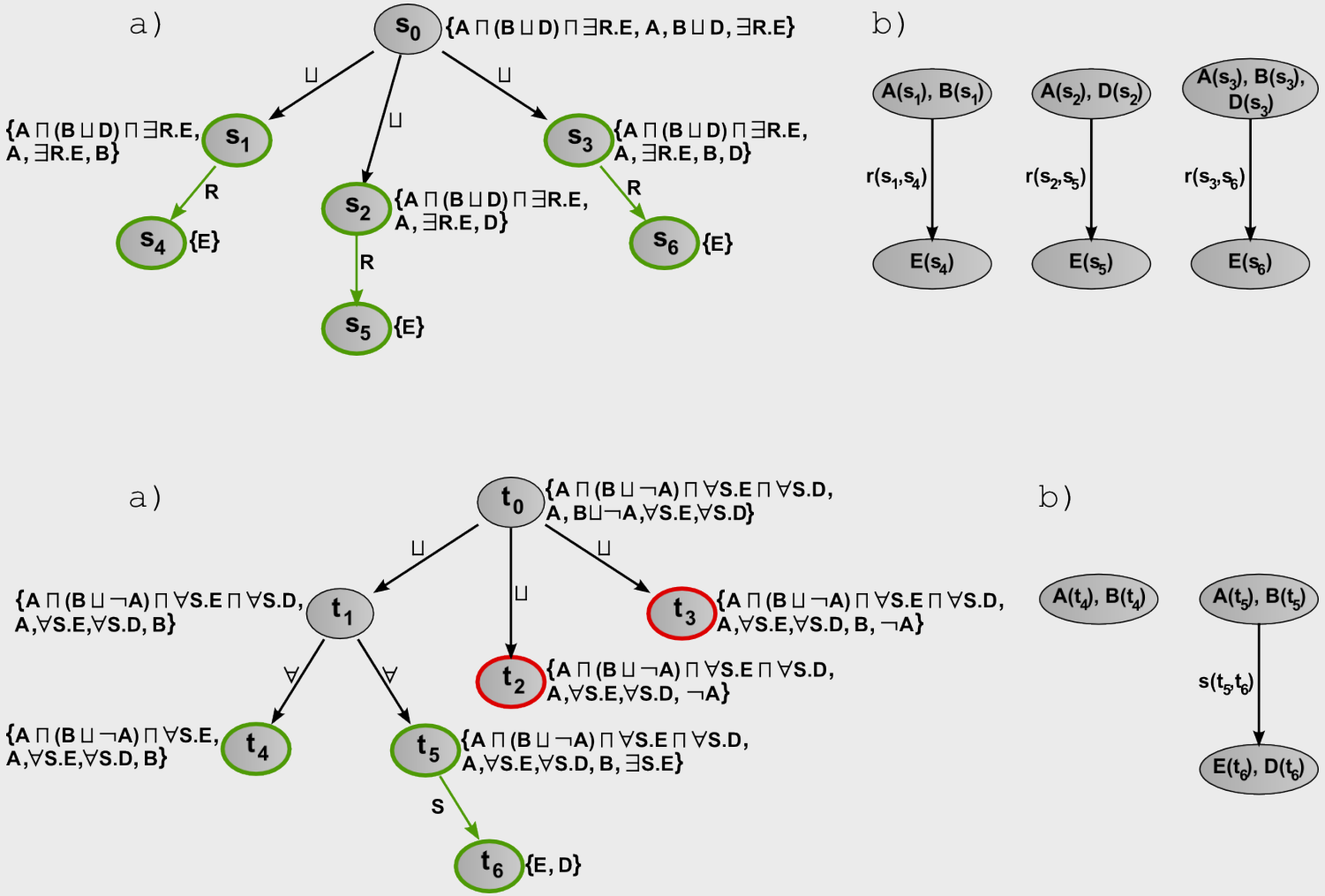
$\mathcal{L}(y) := \mathcal{L}(x)$ (y will then be blocked)

$\mathcal{L}(z) := \mathcal{L}(x) \cup \{\exists R.C\}$.

...

$\geq$ and $\leq$-rules ...

SIM – DL_A: A Simplified Example



SIM – DL_A: Alignment Matrix (for max. Similarity)

a)

Proxy Models of C_s and C_t

C_t \ C_s	s_1	s_2	s_3
t_4	0.67	0.5	0.67
t_5	0.83	0.67	0.63

b)

Assertions of s_1 and t_5

t_5 \ s_1	A	B	$r(s_1,s_4)$
A	1	0.5	-
B	0.5	1	-
$s(t_5,t_6)$	-	-	0.5

c)

Assertions of s_4 and t_6

t_6 \ s_4	E
E	1
D	0

- a) selecting individuals for comparison
- b) selecting assertions for comparison
- c) ...

1 SEMANTICS IN GISCIENCE
- Geoinformation
- Spatial Data Infrastructure
- Semantic Enablement Layer

2 SIMILARITY (REASONING) IN GISCIENCE
- ADL Gazetteer Interface
- SimCat Gazetteer Interface
- SIM-DL Server and Plugin

3 $SIM - DL_A$
- Why $SIM - DL_A$?
- The Similarity Tableau for $\mathcal{SHI}$
- A Simplified Example

4 DISCUSSION & OUTLOOK

DISCUSSION & OUTLOOK

- Support more expressive DL (e.g., cardinality restrictions)
- Integration into existing reasoners (e.g. FaCT, Pellet)
- Approximation and optimization (thresholds –> roles)
- Evaluation of $SIM - DL_A$ (Human Participants Test)

- Conceptualization of and reasoning about events
- Concepts change over time (ecology of concepts)

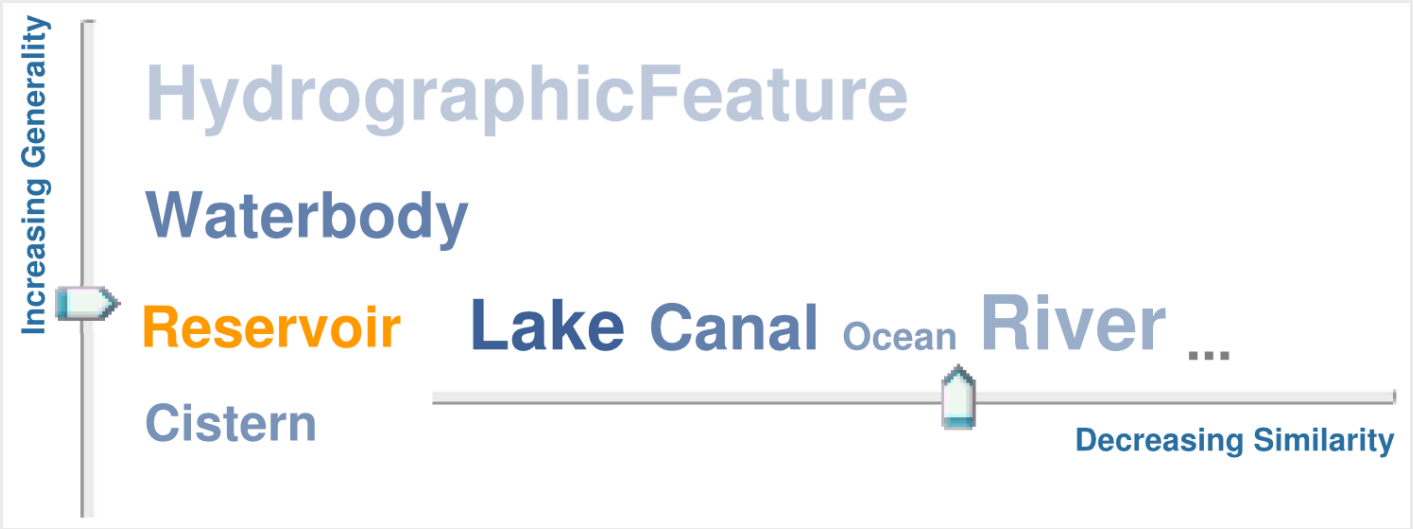
Questions?

Thank you!

- SimCat Project Website: sim-dl.sourceforge.net
- Münster Semantic Interoperability Lab: musil.uni-muenster.de
- International Conference on Geospatial Semantics: www.geosco.org
- Semantic Enablement Community at 52°*North*: 52north.org/semantics

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

NAVIGATION


Increasing Generality
HydrographicFeature
Waterbody
Reservoir
Lake
Canal
Ocean
River
...
Cistern
Decreasing Similarity

QUERY-BY-EXAMPLE

Add reference feature

Selected reference features

Esnagami River - Ontario, CDN

Rideau Canal - Ontario, CDN

Marmion Lake - Ontario, CDN

1

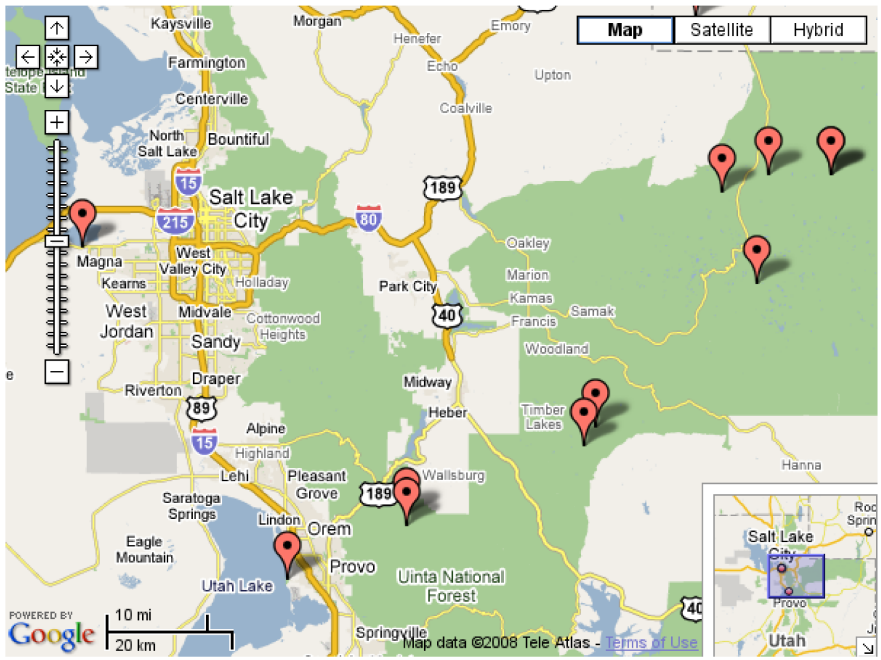
search

similar types

Canal, River, IrrigationCanal, Lake, Reservoir, Tarn

2

Search within this region



14 features found

3

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

<<back 1-10 11-14 next>>

Similarity Reasoning

1. Definition of application area and intended audience
2. Selection of search (query) and target concepts
3. Transformation of concepts to canonical form
4. Definition of an alignment matrix for concept descriptors
5. Application of constructor specific similarity functions
6. Determination of standardized overall similarity
7. Interpretation of the resulting similarity value(s)

COMPLEXITY?

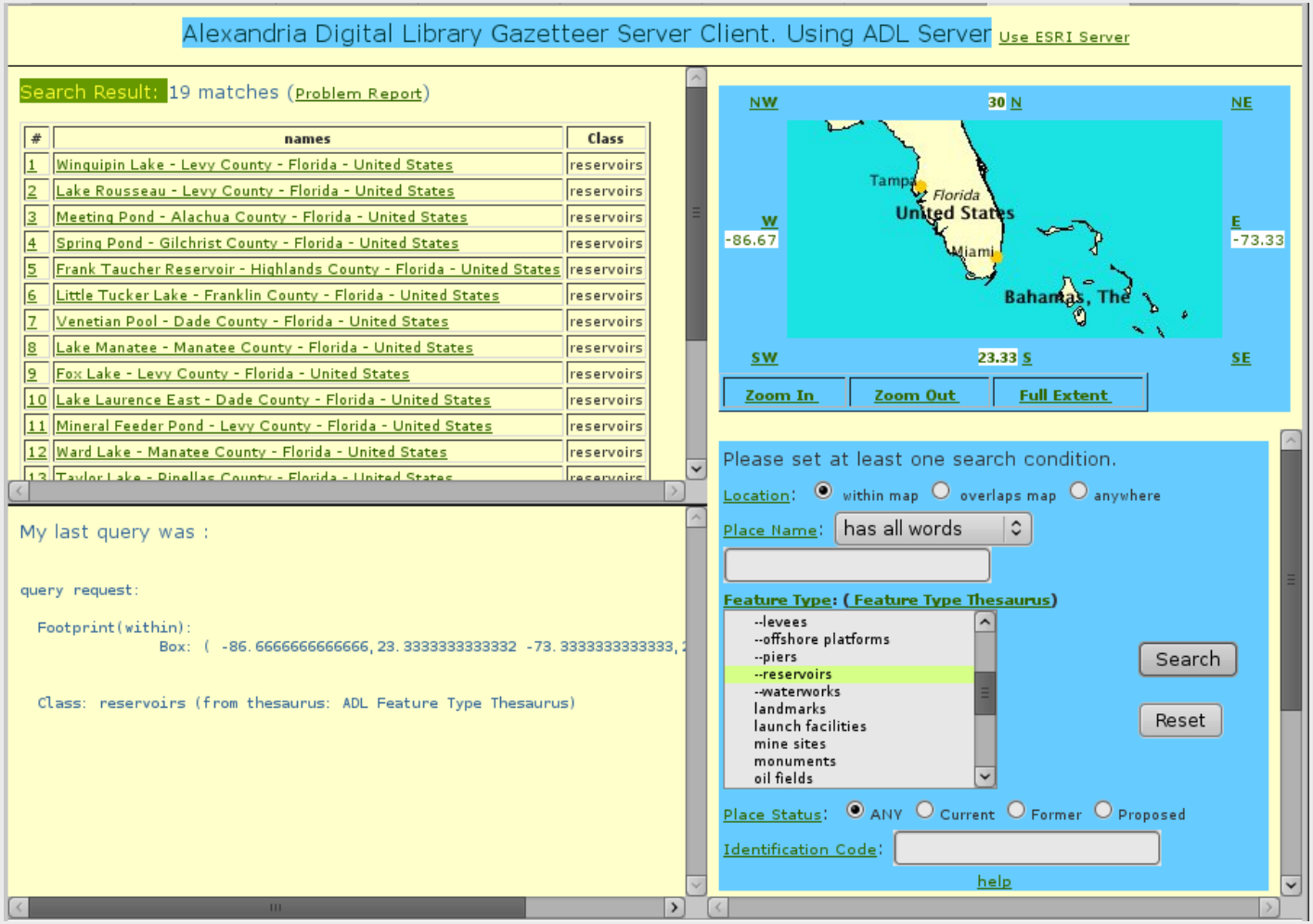
- EXPTIME-complete, but

'[...]This shows that, at the beginning of the new millennium, even an EXPTIME-algorithm is no longer automatically considered to be impractical in the DL community.'

(Baader & Sattler; An Overview of Tableau Algorithms for Description Logics (2000))

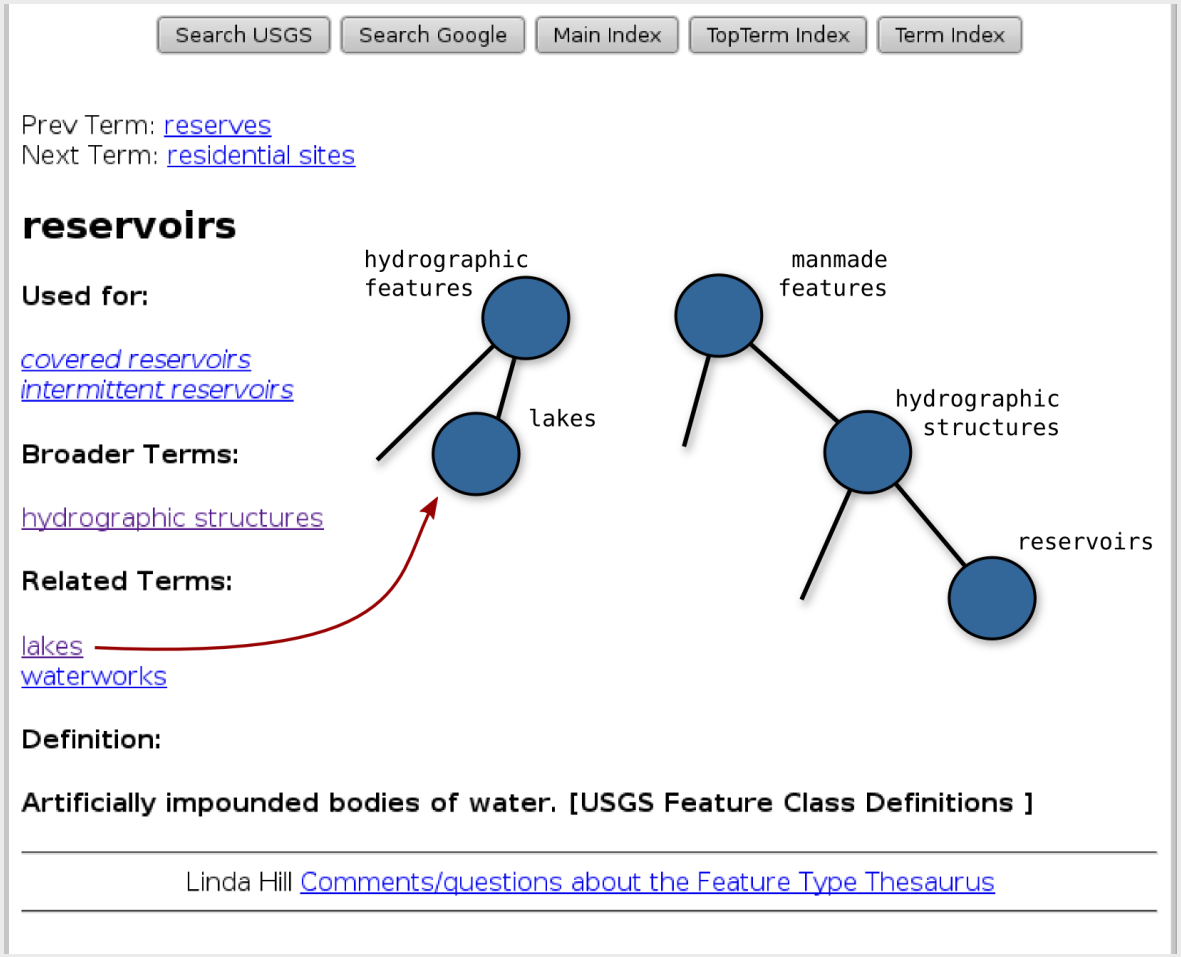
$\longrightarrow$ Optimization and Approximation

THE ADL GAZETTEER WEB INTERFACE & THESAURUS



THE ADL GAZETTEER WEB INTERFACE & THESAURUS

THE ADL GAZETTEER WEB INTERFACE & THESAURUS



THE ADL GAZETTEER WEB INTERFACE & THESAURUS

Search USGS | Search Google | Main Index | TopTerm Index | Term Index

Prev Term: offshore areas
Next Term: oil camps

offshore platforms

Used for:

artificial islands
oil platforms
platforms (offshore)

Broader Terms:

hydrographic structures

Related Terms:

oil fields

Definition:

Structures erected in a sea primarily, but not exclusively, for the extraction of petroleum products. [Adapted from USGS Circ 1048]

Linda Hill Comments/questions about the Feature Type Thesaurus

---

THE ADL GAZETTEER WEB INTERFACE & THESAURUS

hydrographic structures

Used for:

boat landings
boat launches
boat ramps
sluices
water mills

Broader Terms:

manmade features

Narrower Terms:

breakwaters
canals
dam sites
gaging stations
harbors
levees
offshore platforms
piers
reservoirs
waterworks

Related Terms:

hydrographic features
seaplane bases
transportation features
wells

Scope Note:

For constructed bodies of water. For natural water bodies, use 'hydrographic features'.

THE ADL GAZETTEER WEB INTERFACE & THESAURUS

hydrographic structures

Used for:

boat landings
boat launches
boat ramps
sluices
water mills

Broader Terms:

manmade features

Narrower Terms:

breakwaters
canals
dam sites
gaging stations
harbors
levees
offshore platforms
piers
reservoirs
waterworks

Related Terms:

hydrographic features
seaplane bases
transportation features
wells

Scope Note:

For constructed bodies of water. For natural water bodies, use 'hydrographic features'.