ONTIOLOGY BASED DOMAIN SPECIFIC SEARCH OF CROWDSOURCED OPENSTREETMAP DATASET AND WIKI

Prof. Stefan Keller and Michel Ott
Yverdon-les-Bains, 26. October 2012
The Problem: Finding Tags in OSM

- **OpenStreetMap (OSM)**
  - crowdsourced (or volunteered geographic information) mapping project
  - uses an extensible Entity/Attribute/Value schema.
  - ‘tag’ consists of a key and a value
  - p.ex. “viewpoint” is defined as "tourism=viewpoint"

- **Problem**
  - “schema” is changing and growing constantly (curr.~1500 different tags)
  - How to find the most relevant tags?

- **This talk**
  - explains the information retrieval (IR) approach we have chosen
  - a prototype called TagFinder we have implemented
  - future challenges, like constructing thesaurus/ontology
TagFinder in Action

- Search term entered...
  - Schloss

- Search String Analysis
  - Search String: Schloss
  - Detected Language: de
  - Preferred Value: castle

- Find relationships and translations
- Get statistics from Taginfo
- Present result list

<table>
<thead>
<tr>
<th>Count</th>
<th>Key</th>
<th>Value</th>
<th>Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>13659</td>
<td>historic</td>
<td>castle</td>
<td></td>
</tr>
</tbody>
</table>
| 12631  | tiger:county    | New Castle, DE             | [+     ]
| 1045   | tiger:name_base | Castle                     | [+     ]
| 773    | is_in           | New Castle, Delaware, Del., DE, USA | [+     ]
| 760    | gnis:County     | New Castle                 | [+     ]
| 670    | name            | Castle Street              | [+     ]

- Property assertions: castle
  - Related: fortification, fortress, mansion, citadel, palace, OSM
  - Broader: historic
  - Related: stronghold, peel, OSM

- Data property assertions
  - altLabel "Palast"@de
  - altLabel "Chateau"@de
  - altLabel "Prachtbau"@de
  - altLabel "Palais"@de
  - altLabel "Schloss"@de
  - prefLabel "castle"@en
TagFinder: Data Flow

Legend:

→ reading data
→ action

Ontology Based Domain Specific Search...
TagFinder: How it works

Implemented (see slide before):

1. Lookup given input string (any language) in taxonomy/thesaurus
2. If input string found: Get a list of preferred key-value pairs
3. Lookup related terms of each preferred key-value pair for one level
4. Collect statistics from Taginfo for each term
5. Clean tag list
6. Rank according to TagInfo's term statistics
7. Return resulting tag list
Related terms in OSM Wiki pages

Tag: tourism=viewpoint

Available languages
- Deutsch
- English
- Français
- 日本語
- Português do Brasil

Definition
A place where tourists, visitors, hikers might like to visit and take photographs. A place, often high, with a good view of surrounding countryside or notable buildings. Sometimes called a viewpoint, vista point or scenic area/point, lookout or overlook.
Some are identified by signs, many more can be identified from local knowledge or from your own visit and judgement.

This tag identifies a good view, interesting objects or buildings themselves should rather be tagged with tourism=attraction.

Examples

```
tourism=viewpoint
name=Beachy Head
description=Views of the English Channel, France can be seen on a clear day.
```

```
man_made=tower
tower:type=observation
tourism=viewpoint
name=Panorama-Turm
```

See also

```
information=map + map_type=topo
(map indicating the direction to notable landscape features which can be seen from that point)
```

Related terms: view, outlook, viewpoint.

„Related Terms“ (RT) (RT being the common „public“ notion)
Representation of semantics

- **Taxonomy**
  - Subject/term description consisting one relationship: the broader/narrower relationship
  - Builds a hierarchy

- **Thesaurus**
  - Extends a taxonomy
  - Structured vocabulary with "related term" and "preferred term" (inversely: use...) relationships, and "scote note" property

- **Ontology (Topic Map)**
  - Describes "the world" with any set of types, properties, and relationship types.
Example: SKOS encoding of an ontological 'concept'

```
<skos:Concept rdf:about="#castle">
<rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
<skos:prefLabel xml:lang="en">castle</skos:prefLabel>
<skos:topConceptOf rdf:resource="#OSM"/>
<skos:inScheme rdf:resource="#OSM"/>
<skos:altLabel xml:lang="de">Chateau</skos:altLabel>
<skos:altLabel xml:lang="de">Palais</skos:altLabel>
<skos:altLabel xml:lang="de">Palast</skos:altLabel>
<skos:altLabel xml:lang="de">Prachtbau</skos:altLabel>
<skos:altLabel xml:lang="de">Schloss</skos:altLabel>
<skos:broader rdf:resource="#historic"/>
<skos:related rdf:resource="#citadel"/>
<skos:related rdf:resource="#fortification"/>
<skos:related rdf:resource="#fortress"/>
<skos:related rdf:resource="#mansion"/>
<skos:related rdf:resource="#palace"/>
<skos:related rdf:resource="#peel"/>
<skos:related rdf:resource="#stronghold"/>
</skos:Concept>
```

(http://openpoimap.ch/resources/osm_skos.xml)
Possible future work

Linguistic preprocessing:

- Detect search term language
  - Serves also as (spatial) filter to disambiguate terms
  - Example: “Atelier” is either “garage/repair shop” (fr) or an “arts studio” (de)

- Word stemming, fuzzy string matching

- Lookup online word translation list

But the biggest challenge is...
How to extract and crowdsource a taxonomy (thesaurus/ontology) … understandable, and sustainable?

Given Pareto‘s 80/20 rule and besides Related Terms…
THANK YOU!

www.openpoimap.ch/tagfinder

Stefan Keller & Michel Ott
University of Applied Sciences Rapperswil (HSR)
CH-8640 Rapperswil (Switzerland)
sfkeller@hsr.ch
Appendix: From Thesaurus to Ontology...