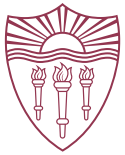


Automatically Constructing Geospatial Feature Taxonomies from *OpenStreetMap* Data

Basel Shbita, Craig A. Knoblock

USC Information Sciences Institute

18th IEEE International Conference on Semantic Computing (ICSC 2024)

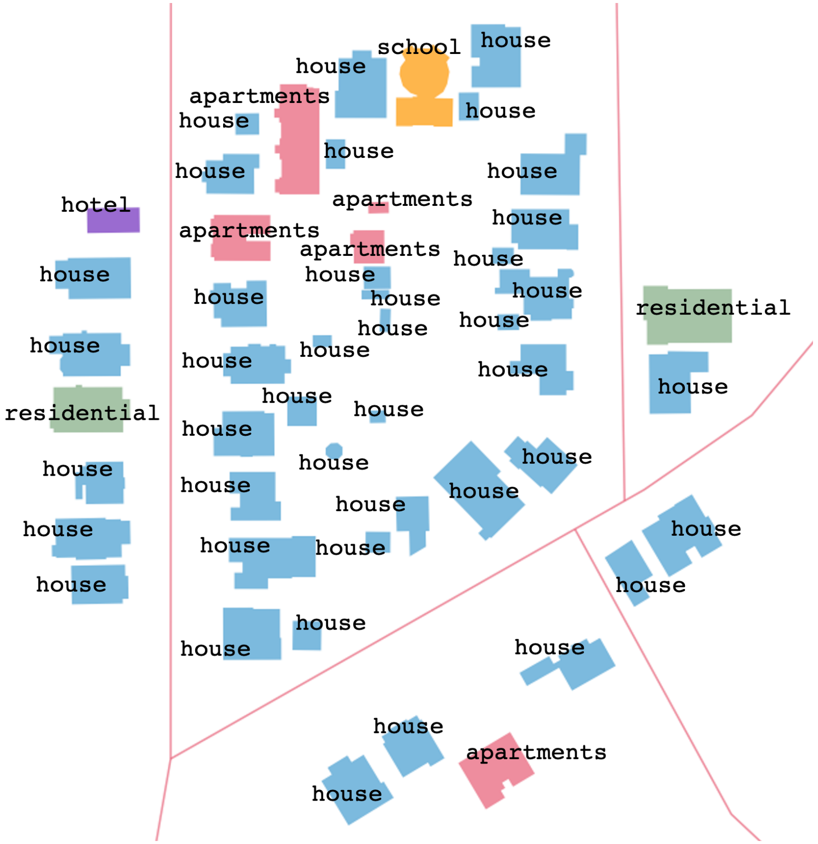
02/06/2024

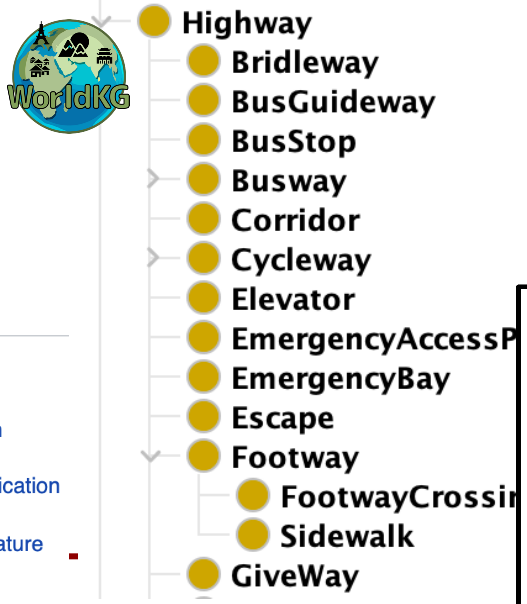
Agenda

- Intro
- Motivation
- Problem
- Approach
- Demo
- Evaluation
- Results & Discussion
- Related Work
- Future Work
- Conclusions

Intro

- accurate & comprehensive characterization of geospatial data in GIS
  - *urban planning, route optimization, navigation systems, remote sensing …*
- structured taxonomy for geospatial features





OSMonto - An Ontology of OpenStreetMap Tags

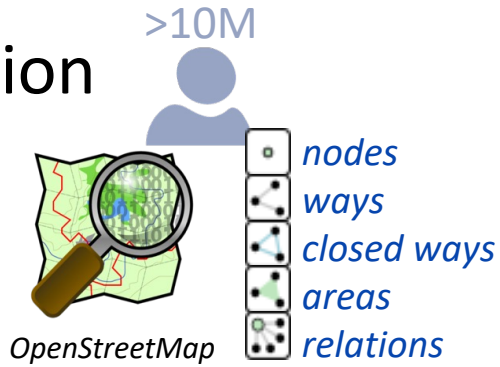
Mihai Codescu*, Gregor Horsinka*, Oliver Kutz**
Till Mossakowski*,**, Rafaela Rau*

* DFKI GmbH Bremen, Germany
** Research Center on Spatial Cognition, SFB/TR 8, University of Bremen, Germany

OUTDATED

Motivation

- *OpenStreetMap* (OSM) = rich source of geographic information
  - VGI (Volunteered Geographic Information)
    - relies on user contributions
  - geometries & attributes of both natural & urban features
  - limited…
    - no standardized taxonomy
    - heterogenous annotations
    - varying-granularity (“how specific”)
    - inconsistent across regions
    - scale
  - can we still make use of this ***noisy*** data?

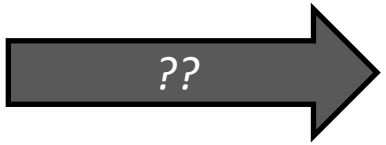


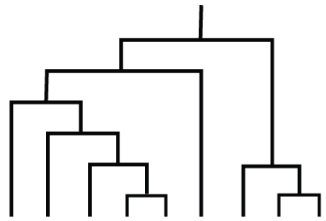
Formalizing the Problem

- How can we establish a comprehensive taxonomy of geospatial features from an unstructured crowdsourced groups of tags, automatically?

"dynamic"

OpenStreetMap data/dump





- Data-driven
  - "application" aware
  - "context" (region) aware
  - automatic

Approach

OpenStreetMap data/dump

```
<way id="232250107" visible="true" vers
2019-05-06T23:22:23Z" user="Enock4seth"
 <nd ref="5058536215"/>
 <nd ref="1797433673"/>
 <nd ref="4992821222"/>
 <tag k="highway" v="tertiary"/>
 <tag k="name" v="Nana Kana Street"/>
</way>
<way id="244376453" visible="true" vers
2015-04-02T14:55:17Z" user="sidneys" ui
 <nd ref="2517024878"/>
 <nd ref="2517024879"/>
 <nd ref="2517024880"/>
 <nd ref="2517024881"/>
 <nd ref="2517024878"/>
 <tag k="building" v="industrial"/>
</way>
<way id="244376454" visible="true" vers
2015-04-02T13:43:25Z" user="sidneys" ui
```

construct base terminology

frequent non-informative

infrequent informative

```
{'apartments',
 'building',
 'driveway',
 'highway',
 'house',
 'residential',
 'service'}
```

count parent-child relations

path frequency

parent	child	counter
building	house	15
highway	service	14
building	residential	33
highway	residential	22
building	apartments	2
service	driveway	5

build taxonomy

conflict resolution

```
OSM
├── building
│   ├── apartments
│   ├── house
│   └── residential__building
└── highway
    ├── residential__highway
    └── service
        └── driveway
```

Approach

pseudo

Algorithm 1: Constructing a lightweight taxonomy.

Data: osmDataset
Result: taxonomyTree

for *entity in osmDataset* **do**
    tagPathsCounter[tagPath]++;

for *(tagPath,count) in tagPathsCounter.sort(order=descending)* **do**
    insert_parent_child_pair(taxonomyTree, parent, child);

return taxonomyTree;

name
maxspeed

{highway=service, service=driveway}
highway—service--driveway

leisure=sauna

highway
residential
building

residential__highway
residential__building

```
OSM
├── building
│   ├── apartments
│   ├── house
│   └── residential__building
└── highway
    ├── residential__highway
    └── service
        └── driveway
```

Demo

Usage

```
usage: generate_taxonomy.py [-h] --input INPUT [--output OUTPUT] [--threshold THRESHOLD] [--blacklist BLACKLIST]

Automatically construct a lightweight taxonomy for geographic features using OpenStreetMap (OSM) data.

optional arguments:
  -h, --help            show this help message and exit
  --input INPUT         OSM dump (xml) input filename.
  --output OUTPUT       Taxonomy tree (json) filename.
  --threshold THRESHOLD
                        Minimum frequency threshold per tag.
  --blacklist BLACKLIST
                        (txt) file with tags to ignore (one per line, as seen on OSM).
```

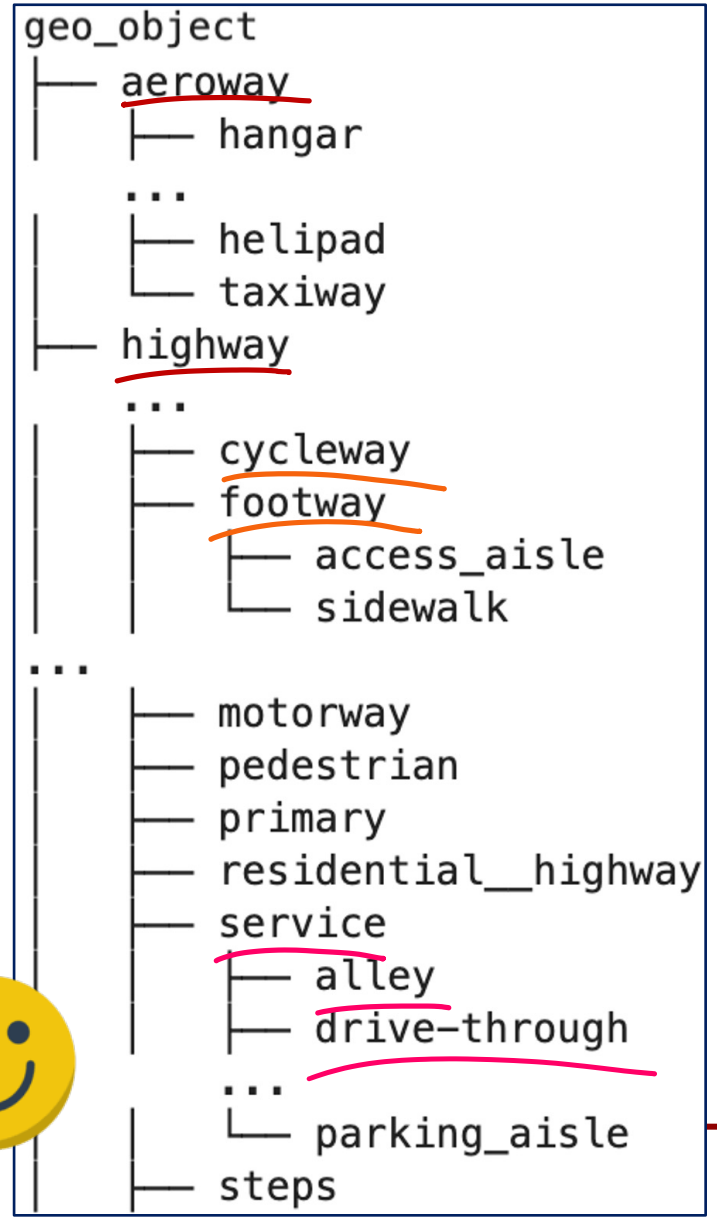
Evaluation

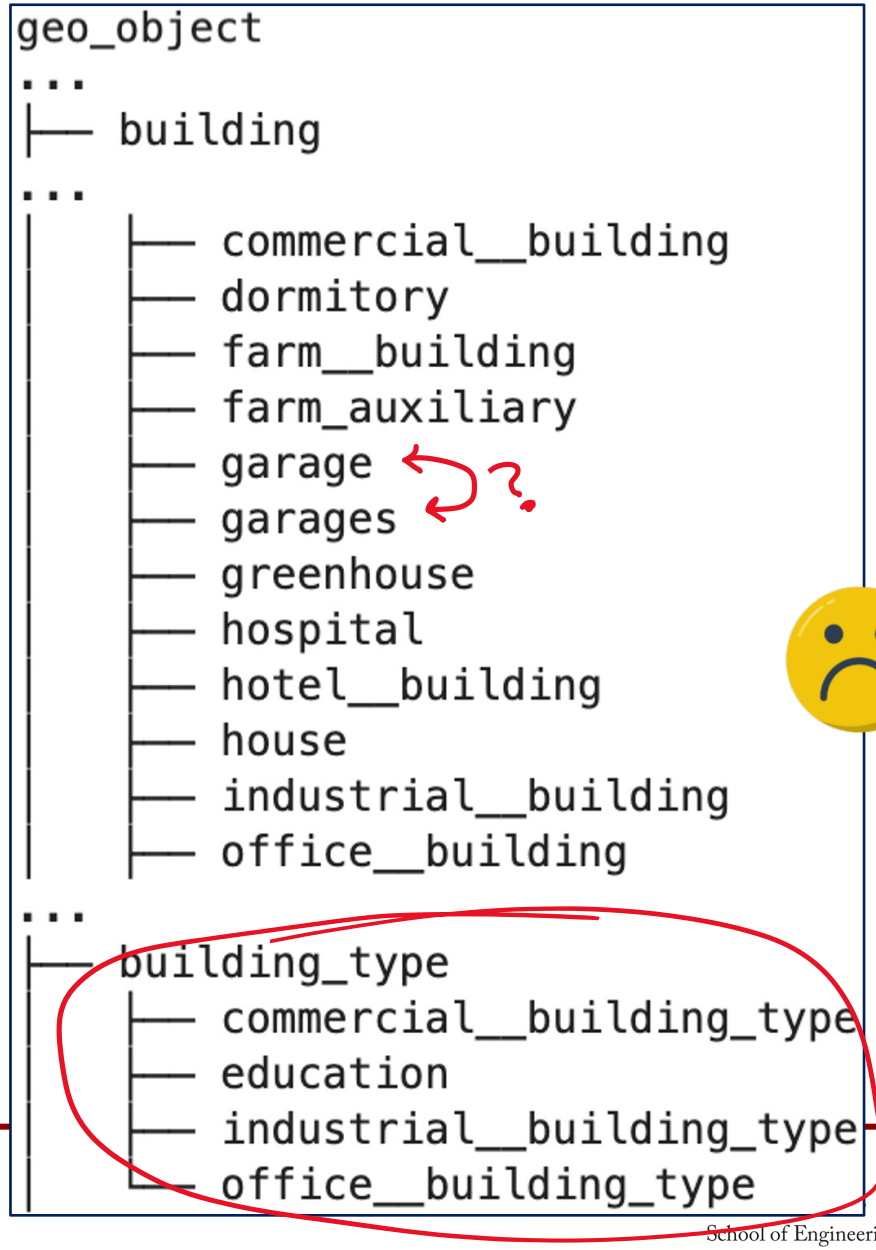
California USA (March 2023)
~150M instances
~10M tagged
1-16 tags (avg 2.3)

Greece (March 2023)
~40M instances
~2M tagged
1-13 tags (avg 2.1)

Results & Discussion

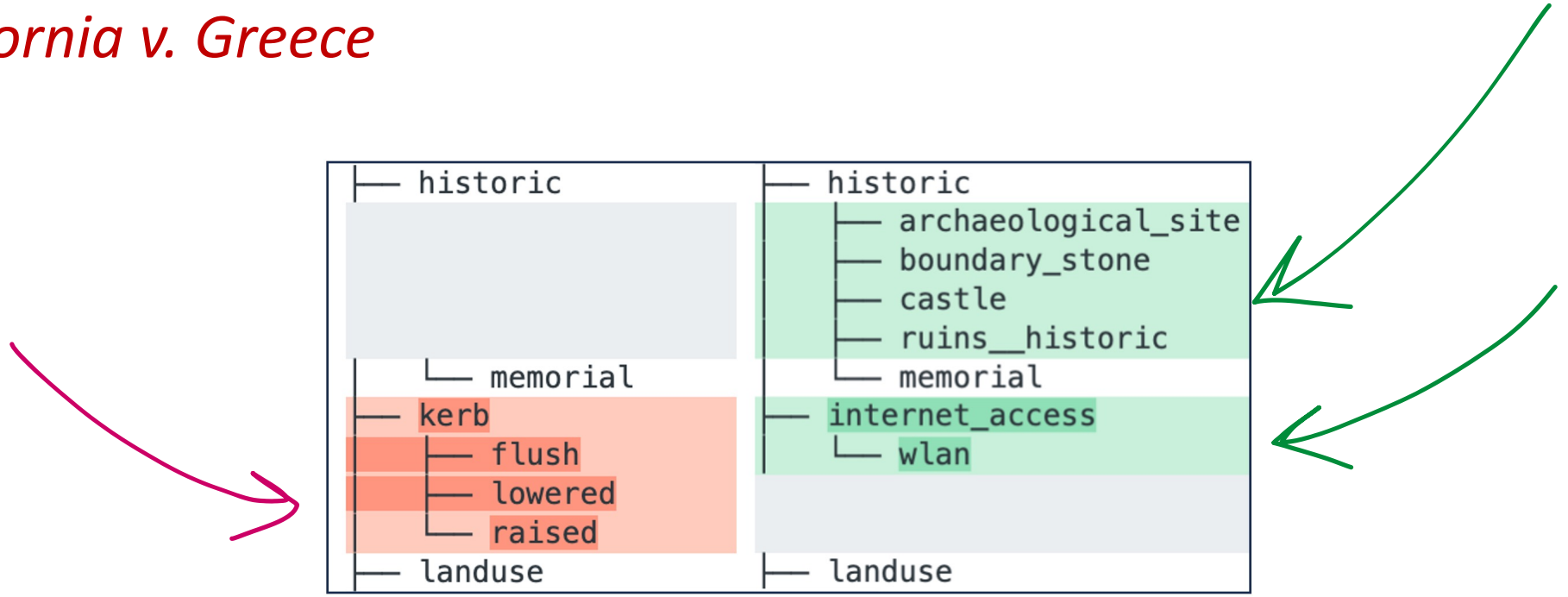
- *California*





Results & Discussion

- *California v. Greece*



Related Work

- Ontologies in Geospatial Data
  - Sun et al. [1]: Three-Level Ontology
    - manual
  - OSMonto [2]: Tag Hierarchies
    - explores tag relationships
  - WorldKG [3]: Geographic Knowledge
    - semantic representation
- Mapping *OSM* tags to Wikidata classes
  - Dsouza et al. [4]: neural architecture for tag-to-class mapping

[1] Sun, K., Zhu, Y., Pan, P., Hou, Z., Wang, D., Li, W. and Song, J., 2019. Geospatial data ontology: the semantic foundation of geospatial data integration and sharing. Big Earth Data, 3(3), pp.269-296.
[2] Codescu, M., Horsinka, G., Kutz, O., Mossakowski, T. and Rau, R., 2011. Osmonto-an ontology of openstreetmap tags. State of the map Europe (SOTM-EU), 2011, pp.23-24.
[3] Dsouza, A., Tempelmeier, N., Yu, R., Gottschalk, S. and Demidova, E., 2021, October. Worldkg: A world-scale geographic knowledge graph. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management (pp. 4475-4484).
[4] Dsouza, A., Tempelmeier, N. and Demidova, E., 2021, September. Towards Neural Schema Alignment for OpenStreetMap and Knowledge Graphs. In International Semantic Web Conference (pp. 56-73)

Future Work

- Scalability
- Technology
  - ML & NLP for ambiguity & reconciliation
- User-centric
  - Incorporate user feedback
  - Tailor to specific applications
- Applications
  - Wider GIS integration

Conclusion

- Unsupervised & automatic approach for constructing lightweight geo-feature taxonomies from *OpenStreetMap* data
  - enhance OSM data usability
  - support data-driven analysis
  - improve geo-feature representation & categorization

- Source code available at:
  - https://github.com/basels/osm-taxonomy

Thank you for listening!

Questions?