Distributed Vocabulary Development with Version Control Systems

Lavdim Halilaj, Steffen Lohmann, Christian Mader, Sören Auer

University of Bonn / Fraunhofer IAIS
Motivation

General GitHub statistics (April 2016):
• > 35 million repositories
• > 14 million users

Well-known ontologies/vocabularies:
• Schema.org
• Description of a Project (DOAP)
• Friend-of-a-Friend (FOAF)
• The Music Ontology
• Edamontology (EDAM)
• Human Disease Ontology
• ...

Our vocabularies:
• Mobivoc
• ScorVoc
• AutomationML
• LiDaKra
• Oddete
• OpenBudget
• ....
Distributed Vocabulary Development

Collaboration Support:
- Governance: roles, permissions, etc.
- Communication: issue tracking, notifications, etc.
- Provenance: revision history, semantic diffs, etc.

Quality Assurance:
- Syntactic Validation: RDF/OWL compliance, etc.
- Semantic Validation: consistency checking, etc.
- Testing: competency questions, etc.

User Experience:
- Documentation: generated HTML, etc.
- Visualization: node-link diagram, etc.
- Editor agnostic: serializations, normalizations, etc.

Vocabulary Deployment:
- Machine accessibility: content negotiation, etc.
- Internationalization: multilinguality, etc.
- Querying: SPARQL endpoint, etc.

→ Partly well-covered by Git-based version control + repository hosting platforms (e.g. GitHub) [Git4Voc16]
Distributed Vocabulary Development

VoCol: An Integrated Environment to Support Version-Controlled Vocabulary Development
VoCol: An Integrated Environment to Support Version-Controlled Vocabulary Development

→ Loose coupling, webhook mechanism, encapsulation via Vagrant & Docker
VoCol: Configuration

Configuration Page

General Info
- Vocabulary Name: Mobivoc
- Domain name: http://vocol.vagrantshare.com
- Web Hook
- Repository info
  - Repository: https://github.com/lavhal-testProj.git
  - Branch Name: master
  - User: lavhal
  - Password: Enter repository password
- Syntax Validation
  - Rapper
  - Jens Riot
- Documentation Generation
  - SchemaOrg
  - Widoco

Additional Services
- Visualization
- Sparql EndPoint
- Syntax Validation Report
- Schema Evolution Report
- Other Branches
- Client Side Hooks
- Turtle Editor
- Use Predefined Queries

Serialization Format
- RdfXML
- NTriples

Save Configuration
VoCol: An Integrated Environment to Support Version-Controlled Vocabulary Development
VoCol: Documentation

ChargingPoint

Definition

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>Ladestation [de]; Ponto de Carregamento [pt]; Punto de Recarga [es]; Pika rimbushese [en]; Charging Point [en]; Point de charge [fr]; Oplaadpunt [nl];</td>
</tr>
<tr>
<td>Comment</td>
<td>Defines the public or semi-public charging points for electric vehicles available worldwide.</td>
</tr>
</tbody>
</table>

Properties

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChargingPointName</td>
<td>Literal</td>
<td>Indicates the name of the charging station</td>
</tr>
<tr>
<td>HasParkingFacility</td>
<td>literal</td>
<td>Indicate whether Filling Station has Parking Facility or not</td>
</tr>
<tr>
<td>accessible</td>
<td>AccessInformation</td>
<td>Access information of the charging point.</td>
</tr>
<tr>
<td>additionalInformation</td>
<td>Literal</td>
<td>Other information about the charging point.</td>
</tr>
<tr>
<td>description</td>
<td>Literal</td>
<td>Description of charging point.</td>
</tr>
</tbody>
</table>
VoCol: An Integrated Environment to Support Version-Controlled Vocabulary Development
VoCol: Querying

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX : <http://butterbur06.iai.uni-bonn.de/>

SELECT DISTINCT *  
WHERE {  
  ?s a ?concept .  
} LIMIT 50
```

<table>
<thead>
<tr>
<th>s</th>
<th>concept</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://eccenca.com/mobivoc/Price">http://eccenca.com/mobivoc/Price</a></td>
<td>rdf:Property</td>
</tr>
<tr>
<td><a href="http://eccenca.com/mobivoc/ElectricalConductivity">http://eccenca.com/mobivoc/ElectricalConductivity</a></td>
<td>owl:DatatypeProperty</td>
</tr>
<tr>
<td><a href="http://eccenca.com/mobivoc/Copper">http://eccenca.com/mobivoc/Copper</a></td>
<td>owl:DatatypeProperty</td>
</tr>
<tr>
<td><a href="http://purl.org/net/mobivoc/Key">http://purl.org/net/mobivoc/Key</a></td>
<td><a href="http://purl.org/net/mobivoc/IdentificationSystem">http://purl.org/net/mobivoc/IdentificationSystem</a></td>
</tr>
<tr>
<td><a href="http://purl.org/net/mobivoc/hasCharger">http://purl.org/net/mobivoc/hasCharger</a></td>
<td>rdf:Property</td>
</tr>
</tbody>
</table>
VoCol: Analytics

Graph Type: BarChart
Query: Statistics - Chart Graphs
SPARQL endpoint: http://butterbur06.iai.uni-bonn.de/fuseki/myDataset/query

![Bar Chart](chart.png)

Number

<table>
<thead>
<tr>
<th>Type</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IdentificationSystem</td>
<td>10</td>
</tr>
<tr>
<td>ParkingFacilityFeature</td>
<td>20</td>
</tr>
<tr>
<td>BookingType</td>
<td>10</td>
</tr>
<tr>
<td>DieselFuel</td>
<td>5</td>
</tr>
<tr>
<td>ParkingFacilityConfiguration</td>
<td>10</td>
</tr>
<tr>
<td>PlugType</td>
<td>20</td>
</tr>
<tr>
<td>ChargingStation</td>
<td>40</td>
</tr>
<tr>
<td>FunctionalProperty</td>
<td>10</td>
</tr>
<tr>
<td>Class</td>
<td>10</td>
</tr>
<tr>
<td>ChargingPoint</td>
<td>20</td>
</tr>
<tr>
<td>ChargingSpeed</td>
<td>5</td>
</tr>
<tr>
<td>ObjectProperty</td>
<td>10</td>
</tr>
<tr>
<td>Vocabary</td>
<td>5</td>
</tr>
<tr>
<td>AccessType</td>
<td>10</td>
</tr>
<tr>
<td>VehicleType</td>
<td>10</td>
</tr>
<tr>
<td>MeansOfTransport</td>
<td>10</td>
</tr>
<tr>
<td>Ontology</td>
<td>10</td>
</tr>
<tr>
<td>DistanceType</td>
<td>30</td>
</tr>
<tr>
<td>Property</td>
<td>60</td>
</tr>
<tr>
<td>Fuel</td>
<td>20</td>
</tr>
<tr>
<td>ParkingFacilityStatus</td>
<td>10</td>
</tr>
</tbody>
</table>
VoCol: Evolution

Comment: Add new properties - Date: 02-04-2016
+ ObjectPropertyRange(<http://butterbur06.iai.uni-bonn.de/payment> <http://butterbur06.iai.uni-bonn.de/Payment>)
+ ObjectPropertyRange(<http://butterbur06.iai.uni-bonn.de/isLocated> <http://butterbur06.iai.uni-bonn.de/ParkingFacilityLocation>)
+ AnnotationAssertion(rdfs:label <http://butterbur06.iai.uni-bonn.de/SpecialParkingRestrictionsInForce> "besondere Parkbeschrankungen in Kraft"@de)
+ AnnotationAssertion(rdfs:label <http://butterbur06.iai.uni-bonn.de/isOwnedBy> "is owned by"@en)
+ DataPropertyRange(<http://butterbur06.iai.uni-bonn.de/threePhasedCurrentAvailable> rdfs:Literal)
Application

• Industrial context (manufacturing company)
• Formally describe the assets of the company
• >10 people (knowledge engineers + domain experts)
• >80 issues, >250 classes, >400 properties, >180 instances
• R2RML mappings: vocabularies & legacy systems
• Queries against the legacy system, visualized results
  + different views on the vocabularies very helpful
Qualitative User Study

- 12 users – different levels of expertise
- Concurrent Think Aloud (CTA) method
- Tasks: define classes, properties, instances
- Commit changes locally, push in remote repository
- Test queries against SPARQL endpoint
- All VoCol functionalities were covered
- Post-study questionnaire (USE test, priorities, pros/cons, suggestions)
Qualitative User Study

+ „Very easy to learn and use“
+ „Very useful and effective support“
+ High usability (all USE scores > 4)
+ Turtle editor, syntax-checking and auto-completion
- More provenance information (author, date, etc.)
- Dynamic configuration
- Recommendation of similar vocabularies
- ...

11/23/2016 VoCol: An Integrated Environment to Support Version-Controlled Vocabulary Development
Conclusions

• Reuse existing VCSs as a core component of vocabulary development
• User-friendly client hiding the complexity of VCSs
• Comprehensive set of integrated services
• Loose coupling, webhooks, Vagrant & Docker
• Flexible architecture, easy to extend
• Further VCS, VoCol-as-service, etc.