Suggesting Mappings from Relational Databases to Ontologies

Jackalopes Group:

Hamza Harkous, Julio Cesar dos Reis, Liang Beet,
Miguel Ceriani, Riccardo Porrini
Hey I’m Charles Bachman, and I invented relational databases!

And I’m Tim Berners-Lee, and I invented the semantic web.
And we met in Kalamaki

Hey Tim

Hey Charles

Hey ESWC Summer School
The Semantic Web is a web of data, in some ways like a global **database**.

Via semantic web, we can improve heterogeneous database integration and **ontology-based data access**.

Aha! And via suggesting ontologies from databases, we can generate huge SW data.
PROBLEMS

Difficulty in selecting appropriate ontologies/vocabularies for reuse.

Need for (semi) automatically suggesting classes and properties suited for a given schema.
POSSIBLE APPROACHES

Creation of a new ontology from an existing database instance

Discovery of mappings between an existing database schema and existing ontologies
We’ll go with your idea Charles!
MANUAL EXPLORATION IN LOV

- Searching for individual key words from the database table fields to retrieve vocabularies
OBJECTIVE

Release Group

id  gid  name  ...

Ranked Concepts
mo:ReleaseGroup
mo:ReleaseEvent
...

Ranked Properties
foaf:name
dc:title
pext:locationName
sioc:name
...

Database

Nodes and Connections
So here are our ideas!
NAÏVE APPROACH

*Release Group*

| id | gid | name | ... |

*Release Group*

N-Gram Similarity

*All Concepts*

*Ranked Concepts*

- mo:ReleaseGroup
- mo:ReleaseEvent
- ...

*All Properties*

N-Gram Similarity

*Ranked Properties*

- foaf:name
- sioc:name
- pext:locationName
- ...

CONToMap
SYNONYMS-ENHANCED APPROACH

Release Group

<table>
<thead>
<tr>
<th>id</th>
<th>gid</th>
<th>name</th>
<th>...</th>
</tr>
</thead>
</table>

N-Gram Similarity

name + Synonyms

Exact Match

All Properties

Ranked Properties

- foaf:name
- sioc:name
- dc:title
- pext:locationName
- ...

...
Refine the property rankings, by trying to group them based on common domains.
CONTEXT-BASED APPROACH -2

Ranked Properties

name
foaf:name
dc:title
...

Exploit the real entities from datasets (e.g. DBpedia)

Ranked Properties

Track length
...
...
...

SUGGESTED EXPERIMENTS

- MusicBrainz case study
- 44 mappings used as gold standard
- Compare the different approaches (e.g., naïve, synonym-enhanced).
- Calculate metrics of precision, recall and f-score.
IMPLEMENTATION

- Datasets:
  - Subset of ontologies from LOV
  - WordNet synsets and hyponyms

- Technologies
  - Local graph store (Jena Fuseki)
  - Java standalone app
  - SQL -> acquiring database schema
  - SPARQL -> querying on LOV
  - Symmetrics Lib -> string similarity
FUTURE WORK

- Experimental evaluation via the MusicBrainz case study
- Refine the approach considering user's selections via interactive interfaces.
Thanks!

Bye guys!

Like he said...