https://vsr.informatik.tu-chemnitz.de/research

TECHNISCHE UNIVERSITÄT CHEMNITZ

Querying the Semantic Web for Concept Identifiers to Annotate Research Datasets

André Langer, Christoph Göpfert and Martin Gaedke

Chemnitz University of Technology, Germany

andre.langer@informatik.tu-chemnitz.de

The Fourteenth International Conference on Advances in Semantic Processing
SEMAPRO 2020
25 – 29 October 2020 - Nice, France

1 Background

Professorship for Distributed and Self-organizing Systems
Chemnitz University of Technology (TUC), Germany

André Langer
PhD Student

Short Résumé

- Born in 1983
- Scholarships from DAAD, SDW and e-fellows
- Research Stay at UCSB, USA in 2006
- Graduated in Media Computer Science at TUC in 2007
- Project Lead for TV Media Online Services until 2016
- Since 2016 PhD Student at TUC, Germany, PhD topic:
- "PIROL: Publishing Interdisciplinary Research Over Linked Data"

- Research Interest:
  Research Data Management, Linked Data, User Interface Experience

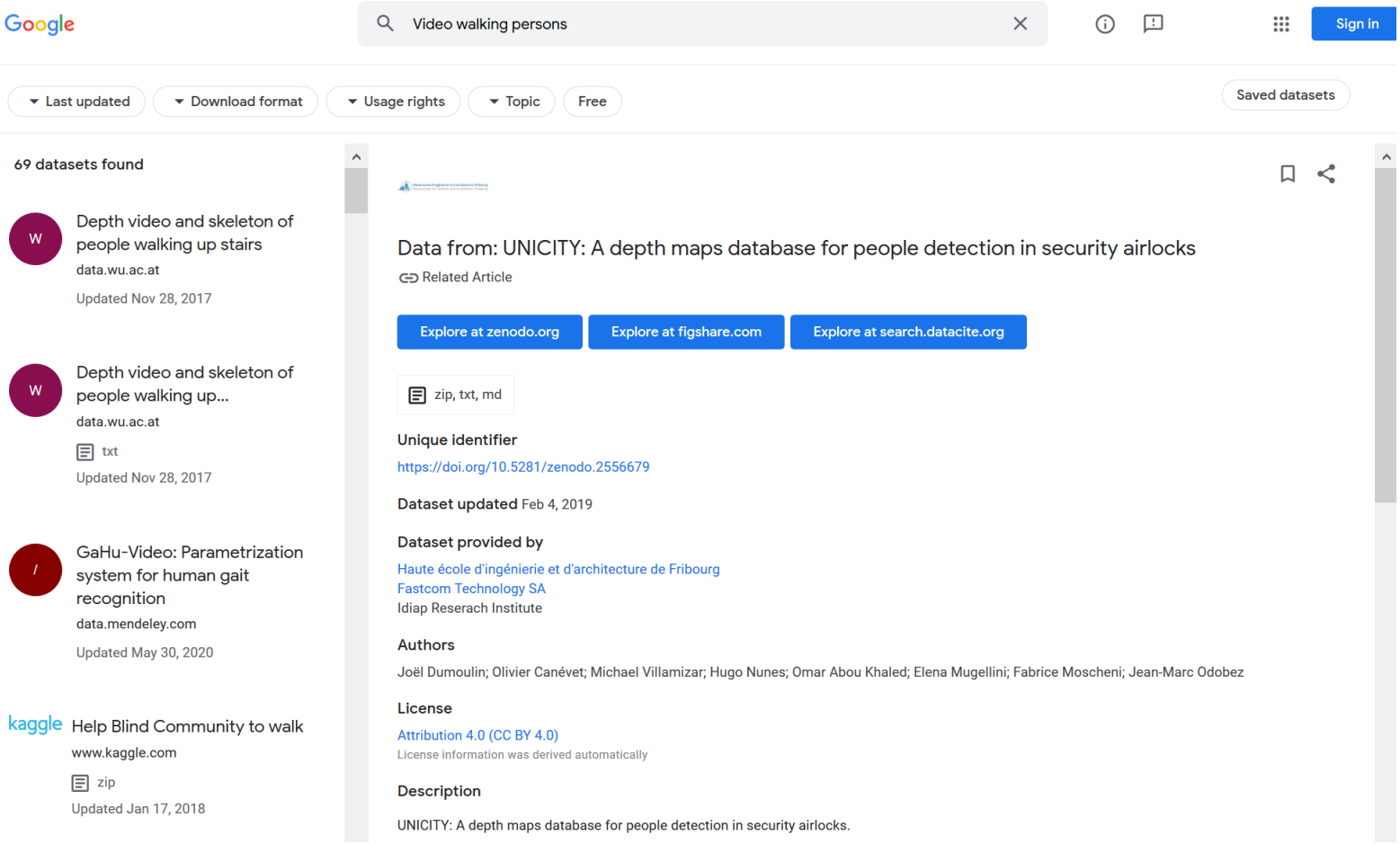
Professorship for Distributed and Self-organizing Systems
Chemnitz University of Technology (TUC), Germany

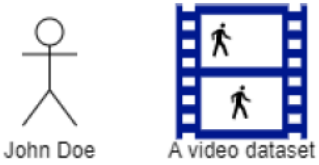
"At VSR, we contribute to enrich the way people live and work in the hyper-connected society by improving human-machine collaboration."

Current Project Context:
DFG Collaborative Research Center „Hybrid Societies“ (2020-2024)

- Establish an Institutional Research Data Repository for Research on Humans Interacting with Embodied Technologies
- Enable Interdisciplinary Discovery and Reuse of Research Artifacts through a common language
- Allow semantic research data annotation based on sophisticated user input web interfaces

2 Problem Description


Google
Video walking persons
Sign in
Last updated
Download format
Usage rights
Topic
Free
Saved datasets
69 datasets found
Depth video and skeleton of people walking up stairs
data.wu.ac.at
Updated Nov 28, 2017
Depth video and skeleton of people walking up...
data.wu.ac.at
txt
Updated Nov 28, 2017
GaHu-Video: Parametrization system for human gait recognition
data.mendeley.com
Updated May 30, 2020
kaggle Help Blind Community to walk
www.kaggle.com
zip
Updated Jan 17, 2018
Data from: UNICITY: A depth maps database for people detection in security airlocks
Related Article
Explore at zenodo.org
Explore at figshare.com
Explore at search.datacite.org
zip, txt, md
Unique identifier
https://doi.org/10.5281/zenodo.2556679
Dataset updated Feb 4, 2019
Dataset provided by
Haute école d'ingénierie et d'architecture de Fribourg
Fastcom Technology SA
Idiap Reserach Institute
Authors
Joël Dumoulin; Olivier Canévet; Michael Villamizar; Hugo Nunes; Omar Abou Khaled; Elena Mugellini; Fabrice Moscheni; Jean-Marc Odobez
License
Attribution 4.0 (CC BY 4.0)
License information was derived automatically
Description
UNICITY: A depth maps database for people detection in security airlocks.



„From Strings to Things“

Persistent Subject Identifier

Some Basic Properties

Unabmiguous Author Identifier

Ambiguity

Missing Typification

No Inference Possibilities

```
{
  "@context": "https://schema.org/",
  "@id": "https://doi.org/10.5281/zenodo.12345678",
  "@type": "Dataset",
  "name": "An elaborate data set on human gait of elderly people"
  "description": "<p>This video dataset comprises 65 recordings of
elderly people moving outside in a parc, recorded with a Sony NEX
50EA.</p>",
  "creator":  [{
     "@id": "https://orcid.org/0000-0001-8672-0508",
     "@type": "Person",
     "name": Doe, John"
   }],
  "datePublished": "2020-07-01",
  "encodingFormat": "video/mp4",
  "keywords": [
    "gait",
    "motion capture",
    "elderly people",
    "male participants",
    "deterioration"
  ],
"researchMethod": "video-based observation",
"license": "CC0 1.0",
  "version": "0.0.1"
}
```

3 Research Question

Which knowledge bases exist that provide relevant concepts to describe research datasets and how can we query them?


4
Approach

Identification of interdisciplinarily relevant concepts

1. Examined established vocabularies for attribute groups
2. Examined UI of established research dataset repositories
3. Examined meta descriptions of existing research datasets

Research area, topic, resource type, (file) format/media type, rights/license, discipline, measurement technique/device, material, audience, demographic characteristics, examined objects, research and evaluation methods, research objective, metrics, measurement characteristics, models

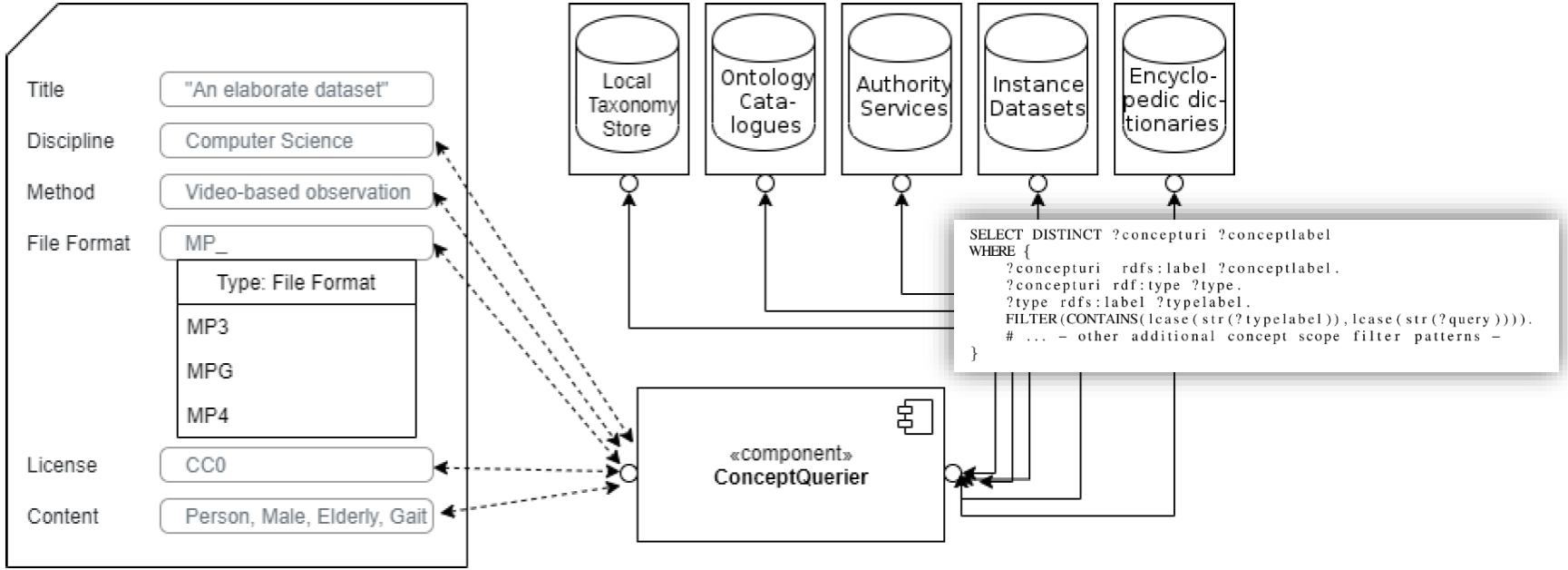
Sources for research data concept identifiers

Ontology Catalogs	Authority Services	Instance Datasets	Other Sources
NCBO BioPortal LOV AberOWL ORR OLS Ontobee IBC AgroPortal SmartCity OC …	EU NALs/Eurovoc LC DNB RAMEAU UNESCO AGROVOC GEMET SSW …	LODCache LOD-a-lot Dbpedia Wikidata BTC YAGO	Static Files Dictionary services Semantic Search Engines

Results

- Appropriate data sources already exist that provide concepts of a certain focus with a concept identifier
- Scattered situation
- Varying data granularity and quality
- Still no or insufficient data providers available for some aspects of interdisciplinary relevance such as methods, devices, objectives
- Variety of representation formats and APIs

Querying Research Concepts



5 Evaluation

Evaluation Strategy

- Prototypical Implementation of *Metadata Input Application to annotate a Research Dataset* with a QueryEngine that dynamically retrieves concepts from external data sources as Proof-of-Concept
- Data and Service Quality Metrics for each data provider group measured based on four example use cases

Evaluation Results

Concept Group	LOV	BioPortal	EuroVoc	Wikidata	DBpedia
Gender	27	37*	4	34	28
License	11	42*	41	435	108
File Format	128	51*	172	4201	432
Research Method	16	149*	0	16	5

Extent of retrieved instances per requested Class Label

Concept Group	LOV	BioPortal	EuroVoc	Wikidata	DBpedia
Gender	1.5s	1.5s*	1.0s	2.7s	0.2s**
License	1.5s	1.5s	1.4s	5.3s	0.2s**
Media Type	1.8s	2.9s	1.0s	5.8s	0.5s**
Research Method	1.5s	3.9s	1.0s	13.2s	0.2s**

Processing Time per requested Class Label in s

6 Conclusion

Contribution

- Analysis of data sources presented that provide concepts and corresponding Linked Data identifiers to describe research datasets
- Implementation of a web-based prototype presented to Dynamically query external services
- Varying service and data quality shown

https://vsr.informatik.tu-chemnitz.de/research

Inspired and Interested?

Andre.Langer@Informatik.TU-Chemnitz.de

@myVSR /myVSR