Vocabulary search

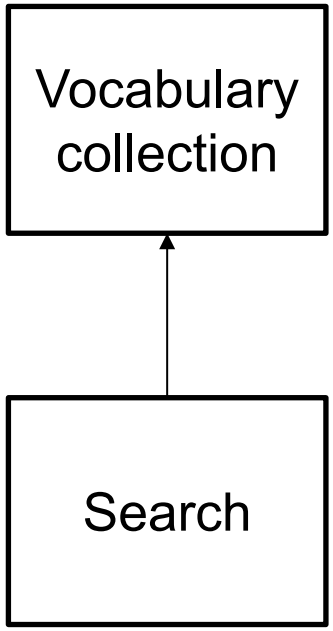
08/10/2019

Niklas Kolbe, PhD student

SnT – Interdisciplinary Centre for Security, Reliability and Trust

Ontology repositories

securityandtrust.lu

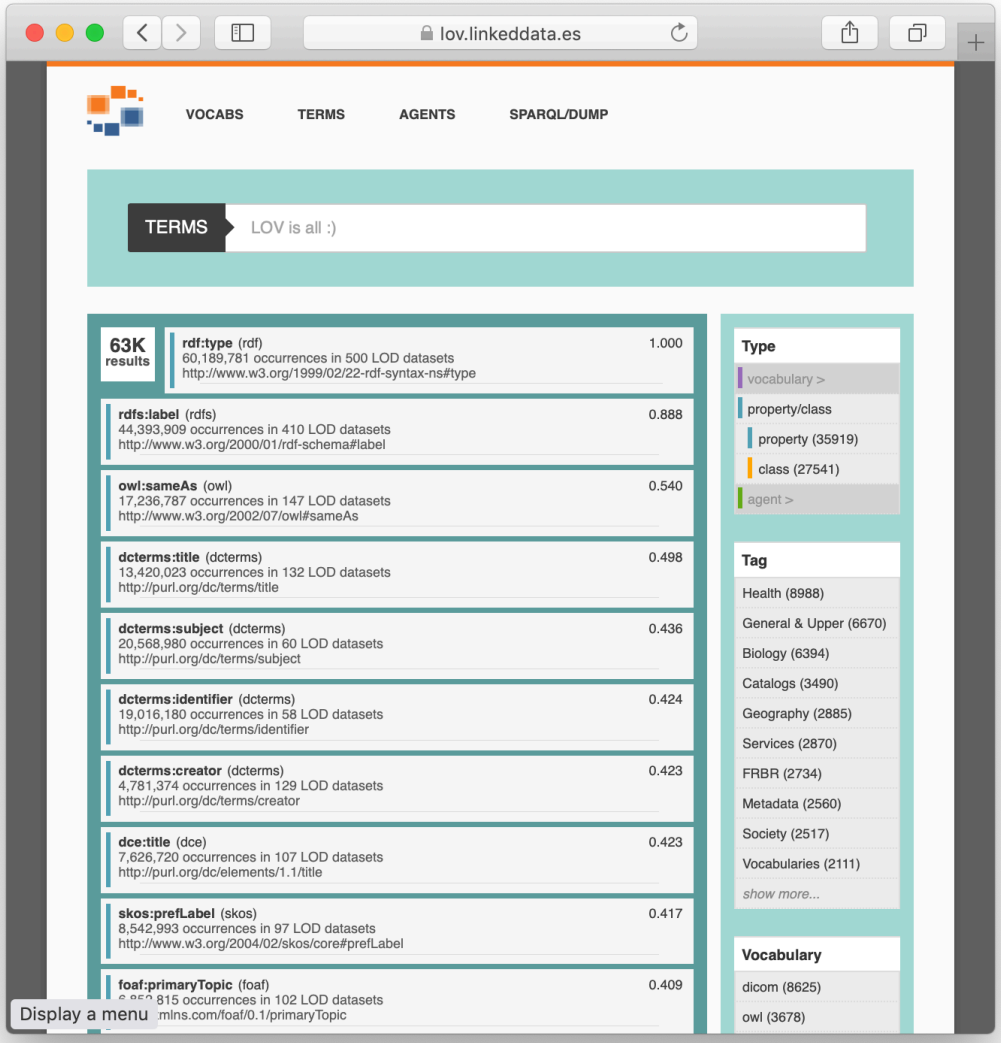


Ontokhoj	2003	[105]	■	✗
OntoSelect	2004	[19]	■	✗
Swoogle	2004	[36]	■	✓
Ontosearch	2005	[170]	■	✗
SWSE + ReConRank	2007	[55]	■	✗
Sindice	2007	[155]	■	✗
Watson	2007	[32]	■	✓
Falcons Concept & Entity Search	2009	[28, 117]	■	✗
VisiNav	2009	[53]	■	✗
WebOWL	2012	[12]	■	✗
LODstats	2012	[8]	■	✓
vocab.cc	2013	[140]	■	✓
OUSAF	2015	[5]	■	✗
Supekar et al.	2004	[147]	■	✗
OntoMetric	2004	[83]	■	✗
Ontology Auditor	2005	[20]	■	✗
OntoQA	2005	[151]	■	✗
Knowledge Zone + TS-ORS	2006	[79, 148]	■	✗
Open Metadata Registry	2006	[60]	■	✓
Ontosearch2	2006	[104]	■	✗
Oyster	2006	[103]	■	✓
OBO Foundry	2007	[139]	■	✓
BioPortal	2009	[100]	■	✓
Cupboard	2009	[33]	■	✗
MMI	2009	[124]	■	✓
Ontobee	2011	[165]	■	✓
BiOSS	2010	[89]	■	✗
Manchester OWL Repository	2014	[90]	■	✓
smartcity .linkeddata.es	2014	[112]	■	✓
LOV	2014	[158]	■	✓
Ontology Lookup Service	2015	[68]	■	✓
Ontohub	2017	[29]	■	✓
(Web)CORE	2006	[25, 38]	■	✗
DWRank	2014	[22, 23]	■	✗
TermPicker	2016	[132]	■	✗
NCBO 2.0	2017	[88]	■	✓
AKTiveRank	2006	[3]	■	✗
(combi)SQORE	2007	[156, 157]	■	✗
LOVR	2015	[142]	■	✓
RecoOn	2016	[24]	■	✓

LOV

securityandtrust.lu

- https://lov.linkeddata.es/dataset/lov/
- 680 ontologies, 43 domains
- 280+ queries / day (UI)
- Ontology + term search
- Curation, metadata, evolution

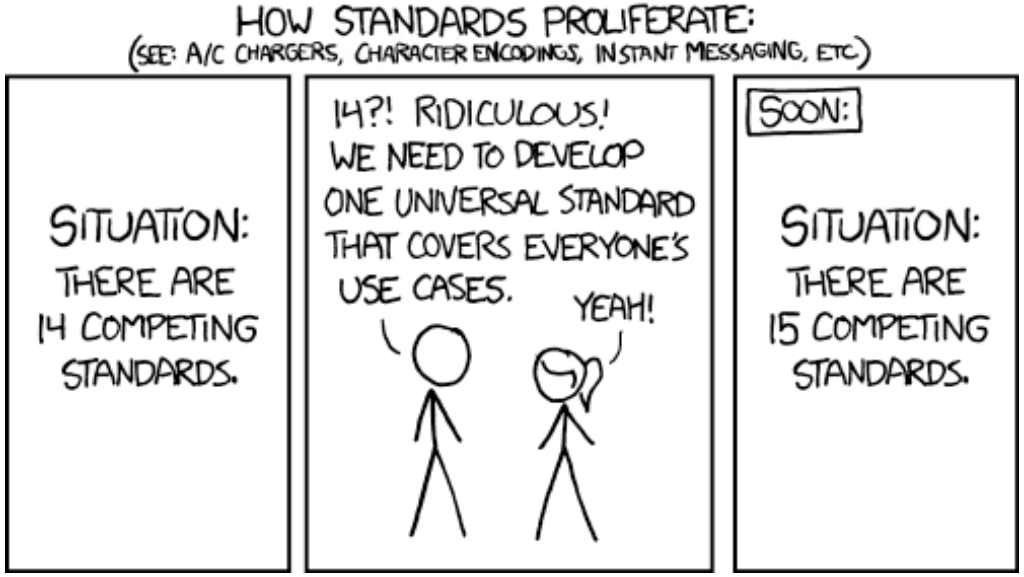


Motivation

securityandtrust.lu

- To help users finding most relevant ontologies and terms through keyword search, e.g., when they are not familiar with a domain

- A good ranking will make it easier to find a suitable ontology for reuse



Relevance

securityandtrust.lu

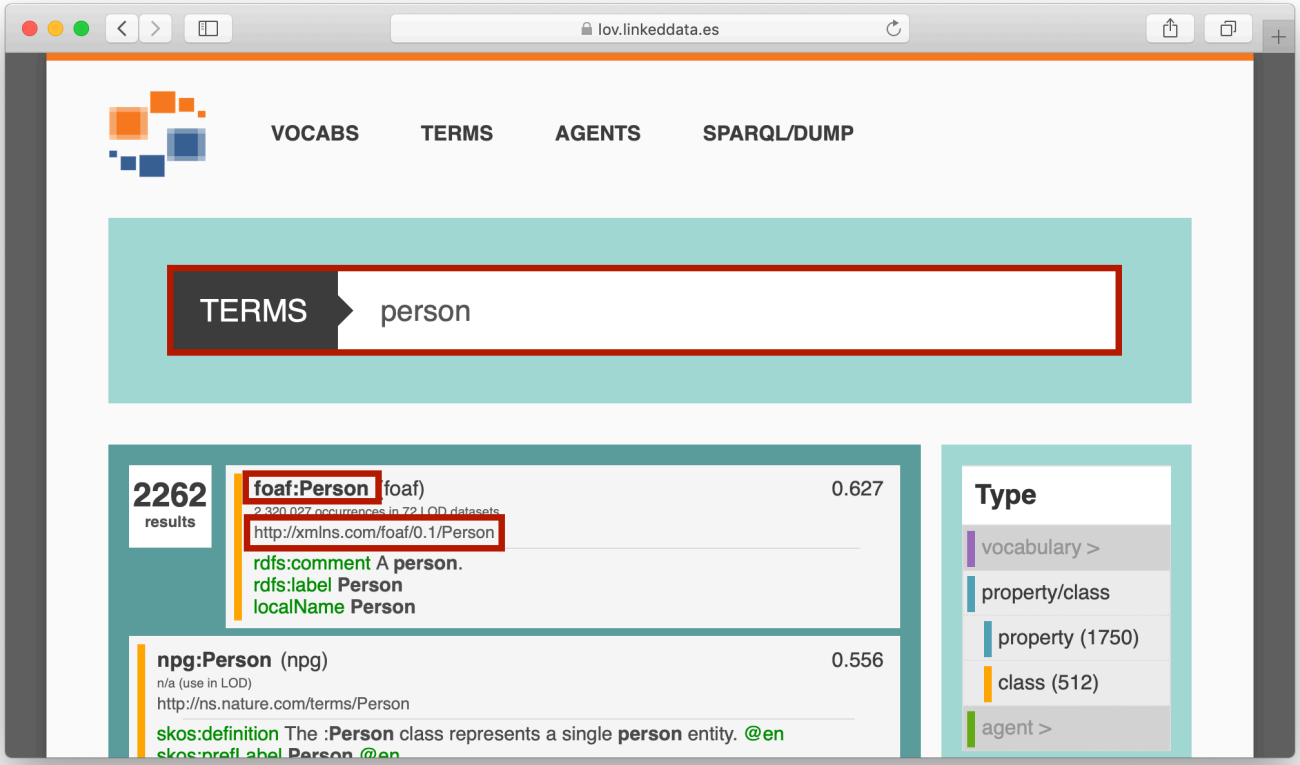
- Expert judgments
  - Costly to scale
  - May contain bias
  - E.g., CBRBench (10 queries)

- Implicit User feedback
  - In the form of observed queries and clicks
  - Contains bias
  - Easy to scale, continuous

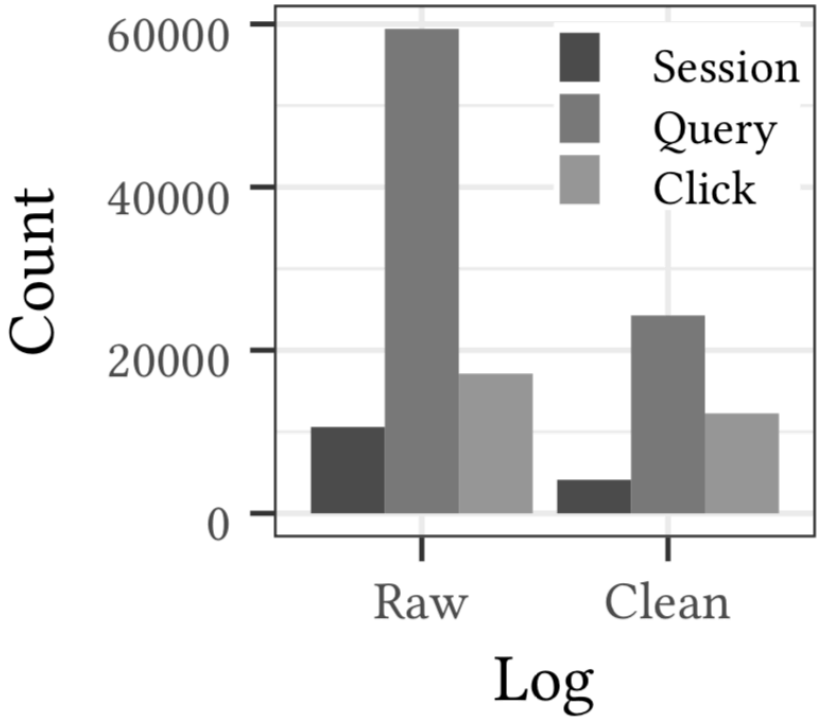
- A large dataset with many relevance judgments allows to learn a ranking model with the optimal combination of several ranking criteria

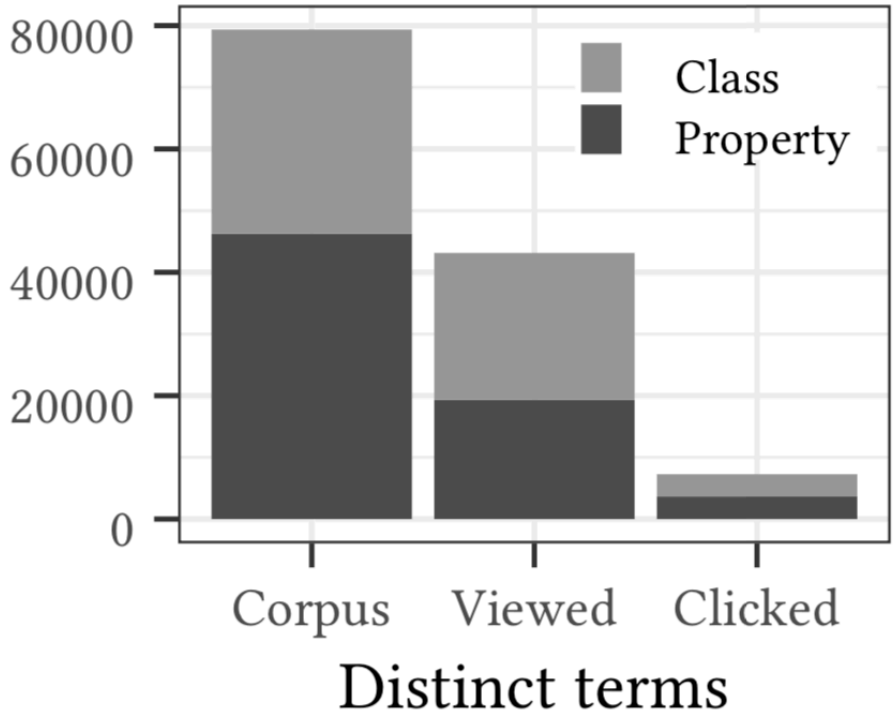
Collecting user feedback in LOV

securityandtrust.lu

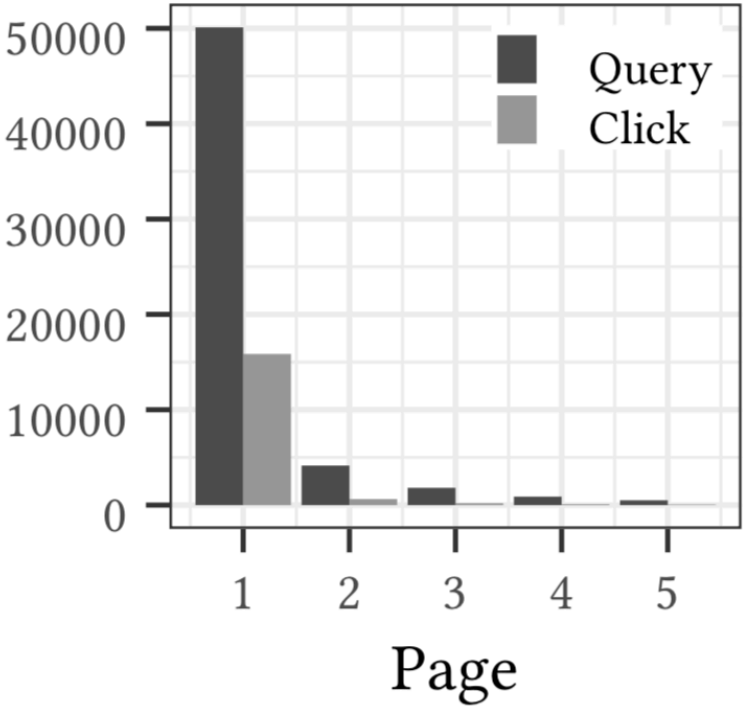


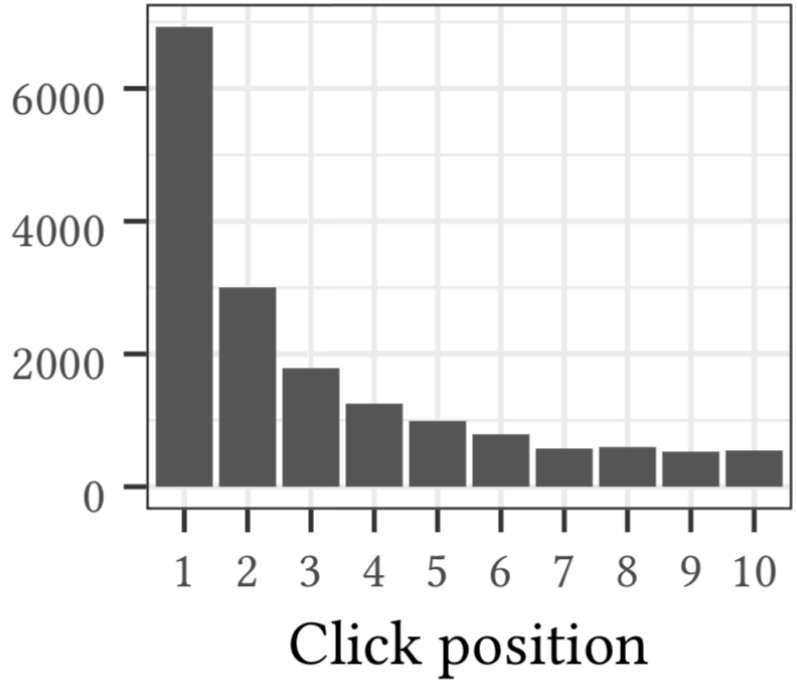
LOV user logs





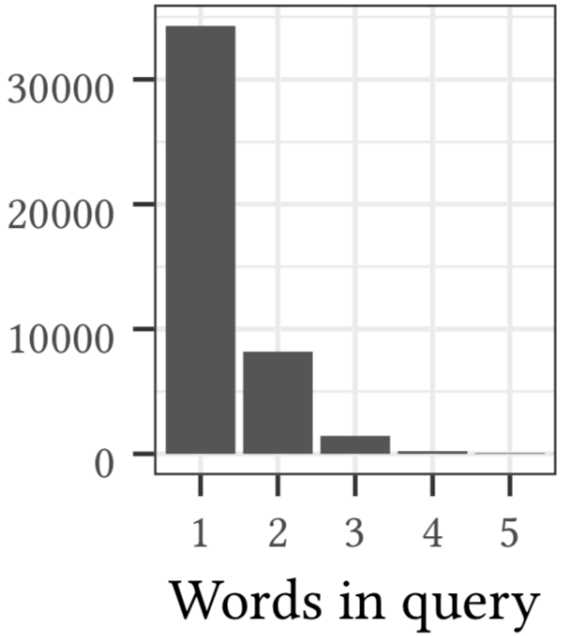
LOV user logs





"The best place to hide a dead body is page 2 of Google"

LOV user logs



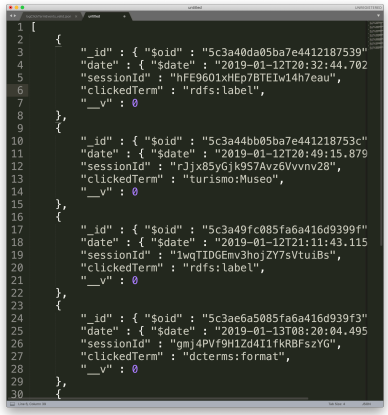
Benchmark (ground truth)

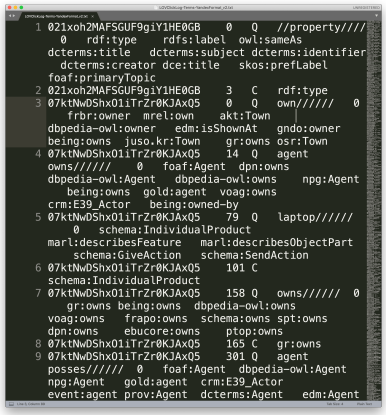
securityandtrust.lu

Server-side logging

Data cleaning (R script)

Learning click model (Python, using PyClick)





LOV

Query and click logs

Cleaned and merged logs

Click probabilities

Benchmark (ground truth)

securityandtrust.lu

- Query – Term – Relevance Score

```
"person","http://xmlns.com/foaf/0.1/Person",4
"person","http://ns.nature.com/terms/Person",0
"person","http://www.bbc.co.uk/ontologies/sport/Person",0
"person","http://schema.org/Person",0
"person","http://purl.org/dc/terms/creator",0
"person","http://purl.org/ontology/af/person",0
"person","http://vocab.getty.edu/ontology#ulan1105_apprentice_of",0
"person","http://d-nb.info/standards/elementset/gnd#dateOfBirth",0
"person","http://dati.san.beniculturali.it/SAN/persona",0
"person","http://purl.org/saws/ontology#marginaliaAddedBy",0
"person","http://d-nb.info/standards/elementset/gnd#NameOfThePerson",0
"person","http://data.vlaanderen.be/ns/persoon#heeftPersoonsrelatie",0
"person","http://d-nb.info/standards/elementset/gnd#firstArtist",0
"person","http://purl.org/dc/elements/1.1/contributor",0
"person","http://vocab.getty.edu/ontology#ulan1310_advised_by",0
"person","http://purl.org/vocab/bio/0.1/Birth",0
"person","http://opendata.aragon.es/def/ei2a#personGender",0
"person","http://vocab.getty.edu/ontology#ulan2841_performer_was",0
"person","http://xmlns.com/foaf/0.1/PersonalProfileDocument",0
"person","http://sparql.cwrc.ca/ontologies/cwrc#FictionalPerson",0
"person","http://rdvocab.info/ElementsGr2/identifierForThePerson",0
"person","http://d-nb.info/standards/elementset/gnd#founder",0
"person","http://www.w3.org/2000/10/swap/pim/contact#preferredURI",0
"person","http://d-nb.info/standards/elementset/gnd#firstComposer",0
"person","http://www.bbc.co.uk/ontologies/coreconcepts/Person",-2
"person","http://www.data-knowledge.org/dk/Person",-2
"person","http://www.aktors.org/ontology/portal#Person",-2
"person","http://purl.org/dc/terms/publisher",-2
```

- 7000+ queries
- 180000+ relevance judgments

Ontology ranking

securityandtrust.lu

- Query match
  - How well does the description match the keyword query?
  - E.g., Lucene search of rdfs:label and rdfs:comment
- Importance
  - What is the standing of the ontology/term in the repository?
  - E.g., how often has it been imported by others?
- Quality
  - what are the characteristics and does it apply to best practices?
  - E.g., existence of labels, consistencies, availability
- Metadata
  - “External” information about the ontologies
  - E.g., how often has an ontology already been used to model data?

Challenges

securityandtrust.lu

- Combination of ontologies
- Personalized search
- Better expression of user information need