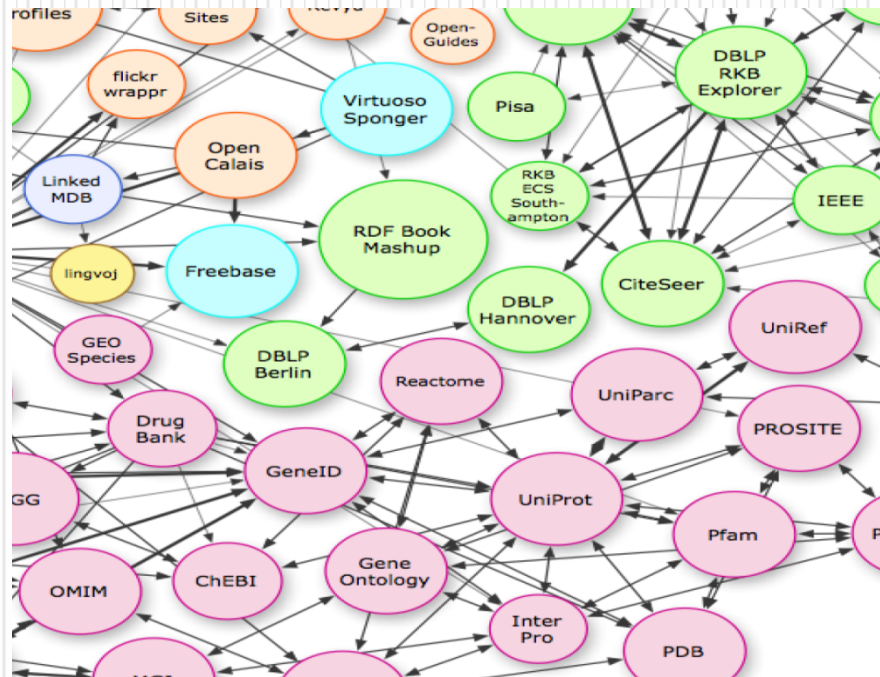


Vicente Palacios, Jorge Morato, Sonia Sánchez, Christos Dimou
University Carlos III of Madrid
LOV Symposium 18th June 2012

Querying linked data: EVALUATION OF SYSTEMS FOR SEMANTIC DOCUMENTS RETRIEVAL



Needs

- A system aimed for retrieving linking metadata vocabularies and other semantic documents depends on:
 - Agreed knowledge representations
 - Unambiguous definitions
 - Availability and accessibility of knowledge
 - Scalability
 - Retrieval capabilities
- Usability

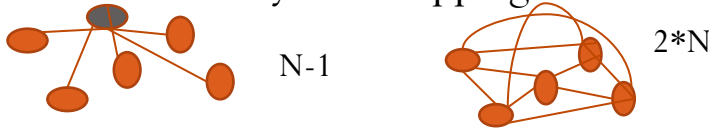
Goal

To analyze a wide range of semantic schema repositories in order to know their suitability for management of semantic documents

Examples

Many linked vocabularies have mistakes in their hierarchy (eg circles), disjoint properties, different classification criteria, granularity, etc.

Scalability problems due to one-by-one mappings between vocabularies

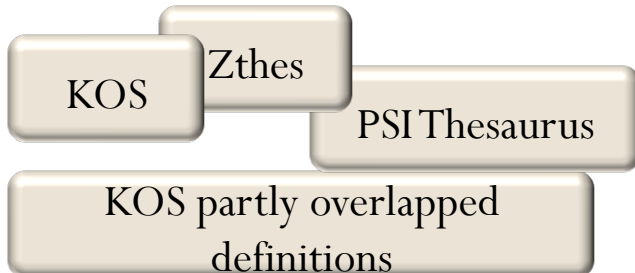


Lack of consensus between similar elements' definitions or languages

Metadata without definition or nonsense elements (e.g. PIM or "dnachecksum" in FOAF)

Need to know metadata identifiers (4.5 mill.) and semantics prior to querying (e.g. [Sparql Virtuoso](#))

...

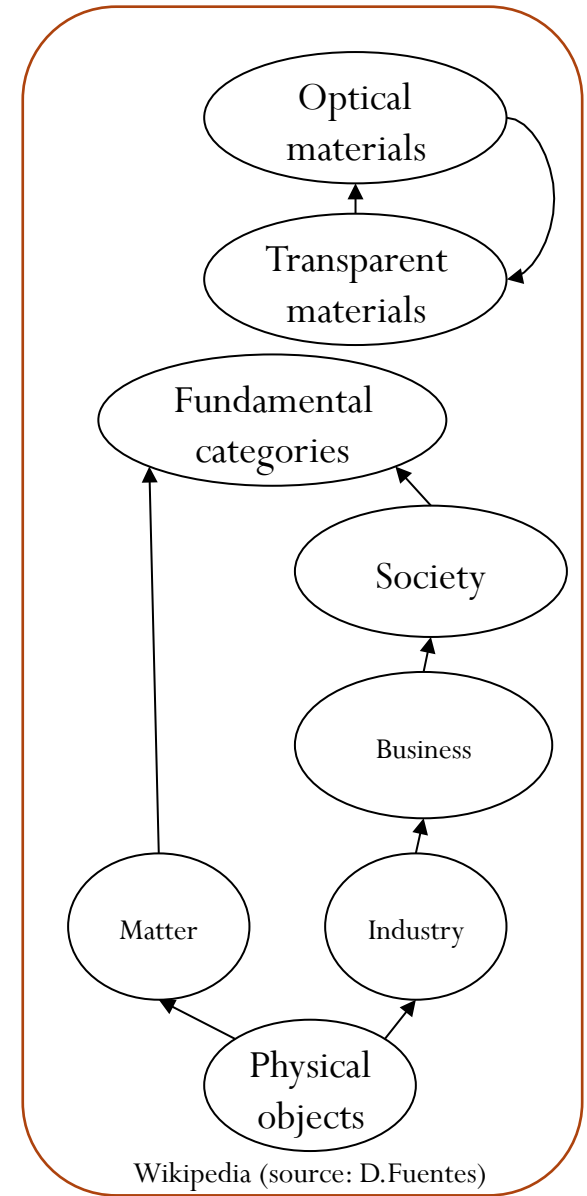


```
SPARQL Explorer for http://dbpedia.org/sparql

SPARQL:
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX dbpedia: <http://dbpedia.org/ontology/>
PREFIX dbo: <http://www.w3.org/2004/02/skos/core#>
PREFIX obo: <http://dbpedia.org/ontology/>

SELECT ?name ?birth ?death ?person WHERE (
  ?person dbo:birthPlace :Berlin .
  ?person dbo:birthDate ?birth .
  ?person foaf:name ?name .
  ?person dbo:deathDate ?death .
  FILTER (?birth < "1900-01-01"^^xsd:date) .
)
```

Results:



Examples: Data Hub

- RelFinder

Transport information from Transport Direct

View Resources (0) Related (0) History

About

Resources (none)

Additional Information

Field	Value
Author	Author not given
Maintainer	Maintainer not given

Google +1 0 Tweet 0 Share 0

Comments

RelFinder

between examples

(1) Leonardo da Vinci
(2) Pablo Picasso

add clear Find Relations

Filter by: relations: (4/4)

length	class	link	conn...
number of objects	num	vi	
1	1/1		
2	3/3		

Leonardo da Vinci

More Info: [dbpedia:arts](#) [en.wikipedia.org](#)

Leonardo di ser Piero da Vinci (April 15, 1452 - 1519) was an Italian Renaissance polymath: painter, sculptor, architect, musician, scientist, mathematician, engineer, inventor,

France

deathPlace

placeOfDe...

Leonardo da Vinci

influencedBy

Roger Ing

influencedBy

influencedBy

Pablo Picasso

nycfacets Smart Open Data Exchange

Search

Log in

Categories Top Datasets Discover Explore Analyze Blog About FAQ Contact

Last visited: Projected Population 2000-2030 - Total By Age Groups (97pn-actf) | Times Square Signage (6bzx-emu4) | Search

Faceted Search

Selected facets

(no facets selected)

Available facets

Categories

- Tabular (8)
- Other (4)
- Text (3)
- Business and Economic... (3)
- Blob (3)
- Facilities and Structures... (1)

Properties

- ▶ GRCUsesQueryCall (14)
- ▶ Modification date (14)

tourism

Find

Sort by Latest article first

Results 1 to 10 of 14

Times Square Signage (6bzx-emu4)

..., pedestrian, entertainment, **tourism**, signage, | NumberOfTags = 15 | NumberOfRows = 184 | Sparseness = 0.55 | NumberOfColumns = 18 | freshnessScore = 7.89 | colinfo = {{{}}}

class="datametable sortable" width="600" {{{}} Position {{{}} Name...

is in category: Other | Tabular

(Show properties)

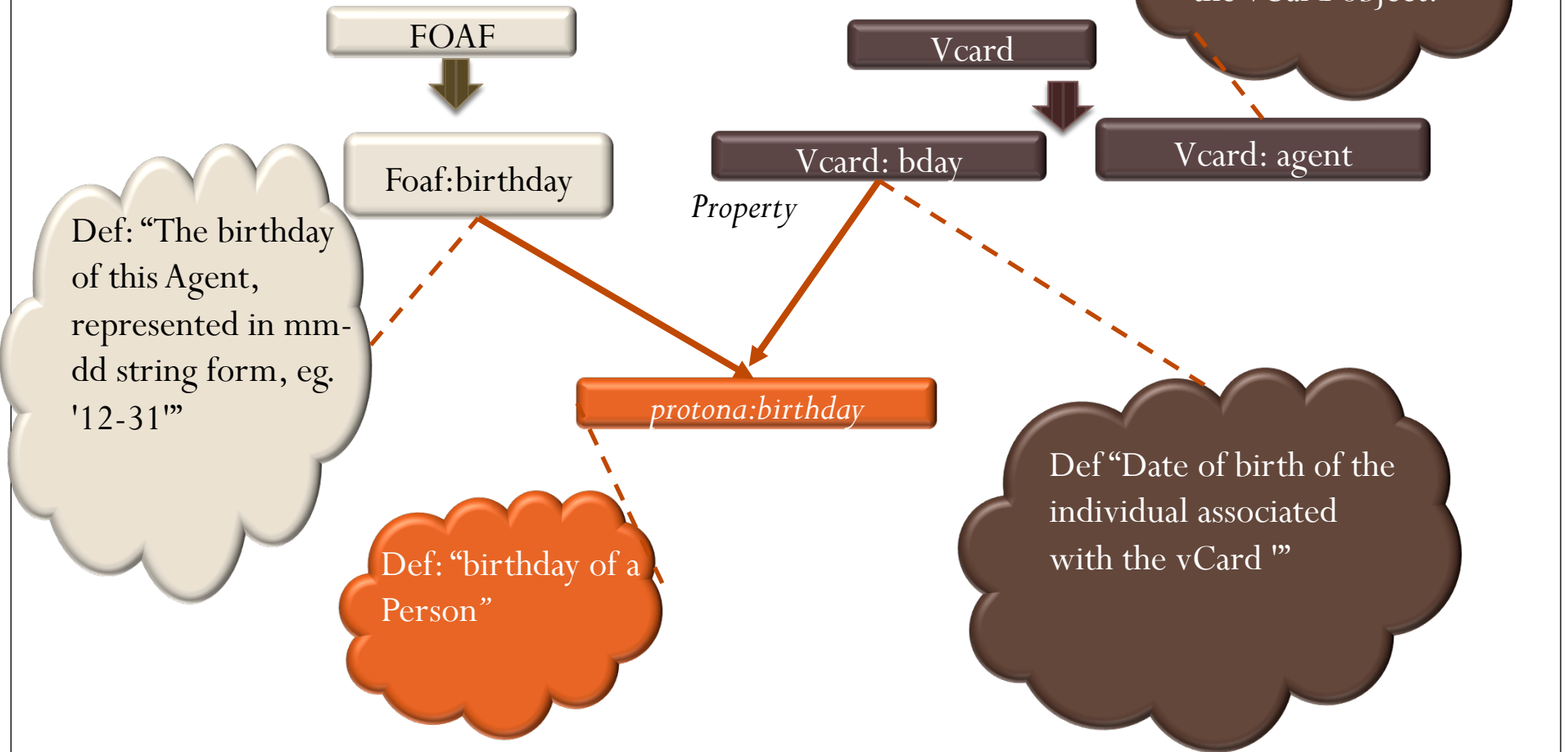
Last change: 2012-06-09 14:44

Times Square Hotels (v8qe-fx6p)

..., manhattan, theater, food, screens, pedestrian, entertainment, **tourism**, | NumberOfTags = 14 | NumberOfRows = 41 | Sparseness = 0.91 | NumberOfColumns = 4 | freshnessScore = 6.79 | colinfo = {

- Data Hub

Examples



Criteria

Resources analysed with Desmet for different criteria:

Schema management	Semantic Management	Query
Interoperability	Disambiguation	Semantic
Formalitation	Multilinguism	Conceptual query
Interactivity	Sinonyms	Contextual query
Semantic Framework	Scope	Document retrieval
	Extensibility	
	Reusability	
	Modifiability	
	Language	

Interoperability	The ability to establish relationships between concepts of different schemas
Formalization	The ability to improve the formalization of a schema
Interactivity	The possibility that the user participates actively in the Schema Management, in accordance to the Web 2.0 guidelines
Disambiguation	The ability to eliminate structural and semantic ambiguity of concepts, in order to facilitate the conceptual retrieval
Semantic Framework	The scope in which the semantic and the conceptual retrieval of concepts is managed.
Multilingualism	Ability to support multiple languages
Synonymy	The ability to solve problems that arise from different concepts with the same meaning
Scope	The domain in which the semantics of the schemas to be managed are defined. It can be either homogeneous or heterogeneous
Extensibility	The ability to expand the representation of the schema semantics
Reusability	The ability to reuse the representation of the schema semantics

Modifiability	The ability to modify the representation of the schema semantics
Language	The ability to represent the language that is use in the formalization of the semantic
Semantic	The ability to express the concrete meaning of a concept in the query process
Conceptual Query	The ability to perform queries, according to the meaning of the concepts
Contextual Query	The ability to obtain results that derive from the existing relationships between concepts
Document Retrieval	The ability to obtain semantic documents that derive from schemas, as well as the schemas themselves

Characteristics and types

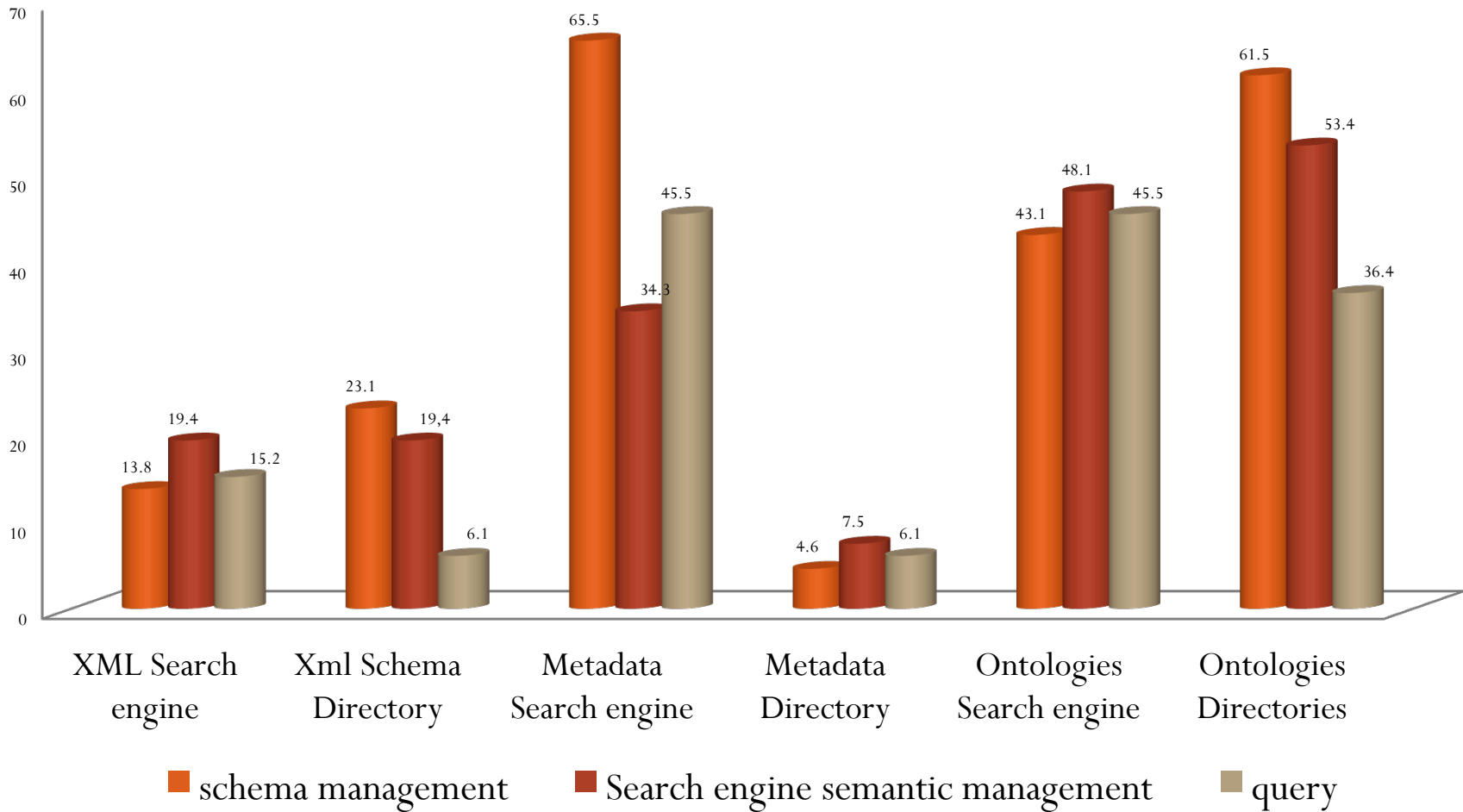
Interoperability	Obligatory/Simple
Formalization	Obligatory/Simple
Interactivity	Desirable/Compound
Disambiguation	Obligatory/Simple
Semantic Framework	Obligatory/Compound
Multilingualism	Desirable /Simple
Synonymy	Obligatory /Compound
Scope	Obligatory/Simple
Extensibility	Desirable /Compound
Reusability	Desirable /Compound
Modifiability	Desirable /Compound
Language	Optional/Compound
Semantic	Obligatory/Simple
Conceptual Query	Obligatory /Compound
Contextual Query	Obligatory /Compound
Document Retrieval	Optional /Simple

Resources evaluated

- Ontology Search Engines (16) *Ontaria, SemanticWeb Search, OwlSeek, SchemaWeb, Swoogle, Falcons, Sindice, Watson, WebKB, Simile, OWLIR, Swotti, Evri, Wolfram-Alpha, Google Squared, MSN Bing*
- Search Engines for XML documents (7) *Quizx, XML-Server de MarkLogic, Xindice, Repository, IxiaSoft, X3 XML Search Engine, XIRCUS*
- Search Engines for metadata (9) *SemanticWebSearch, HotMeta, Semantic Blogging Demonstrator, Learning About Learning Objects, Schemas, Open Metadata Registry, Department Defense Metadata Registry, US Environmental Data Registry, NBII Clearinghouse Search*
- Ontology Directories (7) *SemWebCentral, DAML Ontology Library, Library of Ontologies, WebOnto ontology library, WebODE ontology library, Open Biological and Biomedical Ontologies, NCBO Bioportal*
- XML Schema Directories (3) *_XML.ORG Registry, XMLpitstop, ZVON*
- Metadata Directories (2) *UKOLN Metadata Resources, Public Subject Indicators of Topic Maps*

Characteristic	XML Engines	Schema Directories	Metadata Engines	Metadata Directories	Ontology Engines	Ontology Directories
Interoperability	NO	NO	YES	NO	YES	YES
Formalization	NO	NO	NO	NO	NO	NO
Interactivity	Medium	High	High	Low	Low	High
Disambiguation	NO	NO	NO	NO	NO	NO
Semantic Framework	None	None	Medium	None	Medium	Medium
Multilingualism	NO	NO	NO	NO	NO	NO
Synonymy	None	None	Medium	None	Medium	Medium
Scope	High	High	Low	Low	High	High
Extensibility	Low	Low	Medium	Low	Medium	Medium
Reusability	None	None	None	None	None	Medium
Modifiability	None	None	Medium	None	Medium	Medium
Language	Medium	Medium	Medium	Medium	High	High
Semantics	NO	NO	NO	NO	NO	NO
Conceptual Query	Low	Low	Medium	Low	Medium	Medium
Contextual Query	None	None	Medium	None	Medium	Medium
Document Retrieval	YES	NO	YES	NO	YES	NO

Results



Conclusions

- In the analysis we have selected different kinds of semantic repositories, but the evaluation has reported poor results.
- Most of the analysed systems to manage semantic data in a linked data context have a limited application domain, depend on ill-defined vocabularies (without a design plan to improve the vocabularies in an interoperable way), and have a lack of usability
- In order to handle reuse of semantic documents it is necessary an extensible frame able to include unambiguous definitions, understandable for users and accessible over the Web.