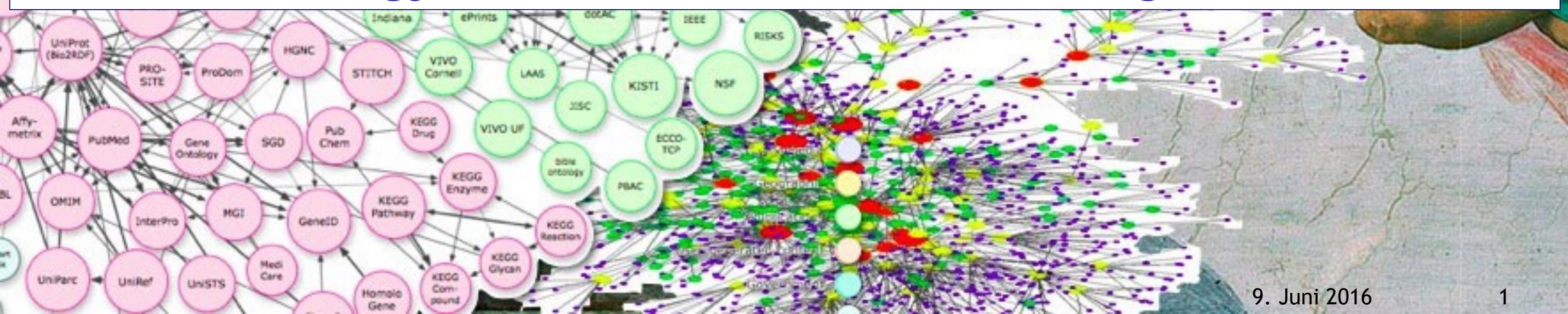


**Vaibhav Kasturia**  
[vbh18kas@gmail.com](mailto:vbh18kas@gmail.com)

**Supervisor: Dr. Stefan Dietze**  
[dietze@i3s.de](mailto:dietze@i3s.de)



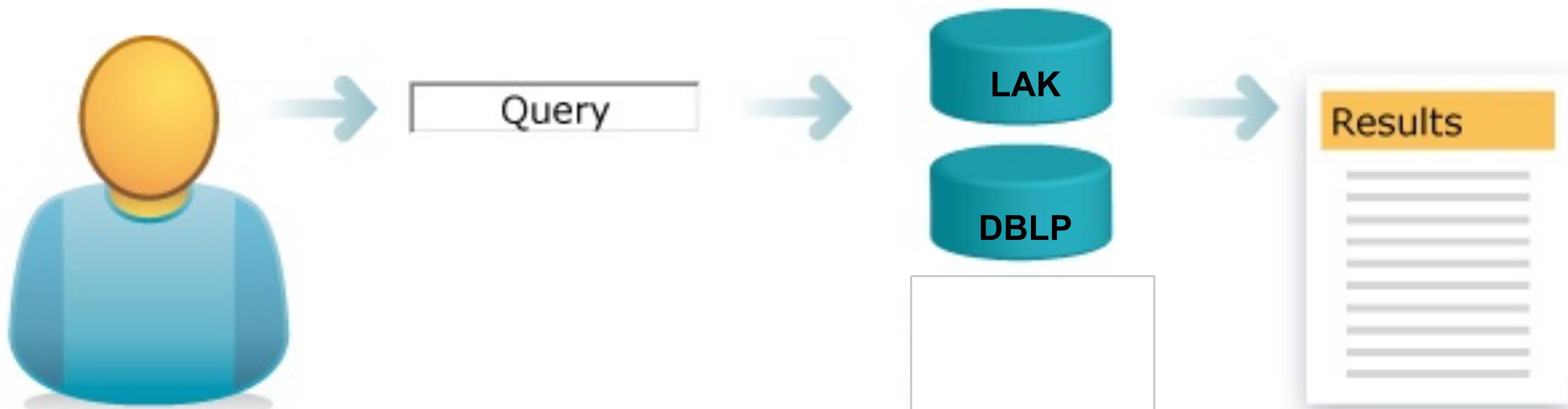
## Outline

- First Approach: Federated Query
- Schema
- Querying with SPARQL
- Second Approach
- Parsing XML File and Generating HashMap
- Coreference Resolution
- Results
- Evaluation
- Conclusion



# First Approach: Federated Query

## *Federated Search*



## Schema

- LAK Dataset
  - Endpoint: <http://meco.l3s.uni-hannover.de:8890/sparql>
  - Default Dataset Name(Graph IRI): <http://lak-conference.org/papers>

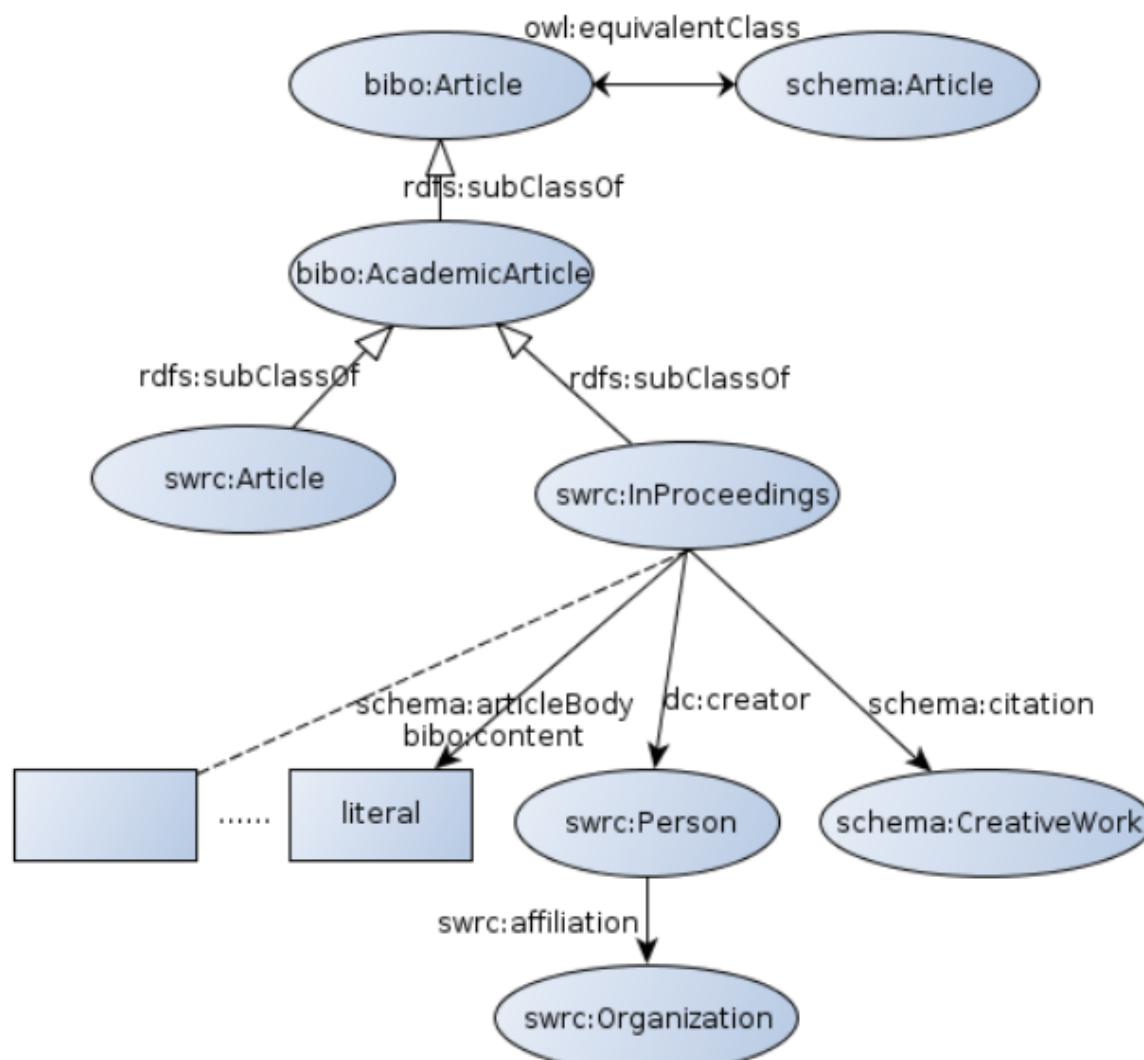


**Learning Analytics and Knowledge (LAK)**

# Schema

- LAK Dataset

→ Schema: <http://lak.linkededucation.org/schema/lak.rdf>



Concept	ofType	#
Reference	schema:CreativeWork	7885
Author	swrc:Person	1214
Conference Paper	swrc:InProceedings	697
Organization	swrc:Organization	365
Journal Paper	swrc:Article	45
Conference Proceedings	swrc:Proceedings	15
Journal Issue	bibo:Issue	9
Journal	bibo:Journal	2

Entity Population

## Key Classes and Properties

Facilitating Scientometrics in Learning Analytics and Educational Data Mining - the LAK Dataset, Dietze, S., Talbi, D., D'Aquin, M., Semantic Web Journal, 2015.



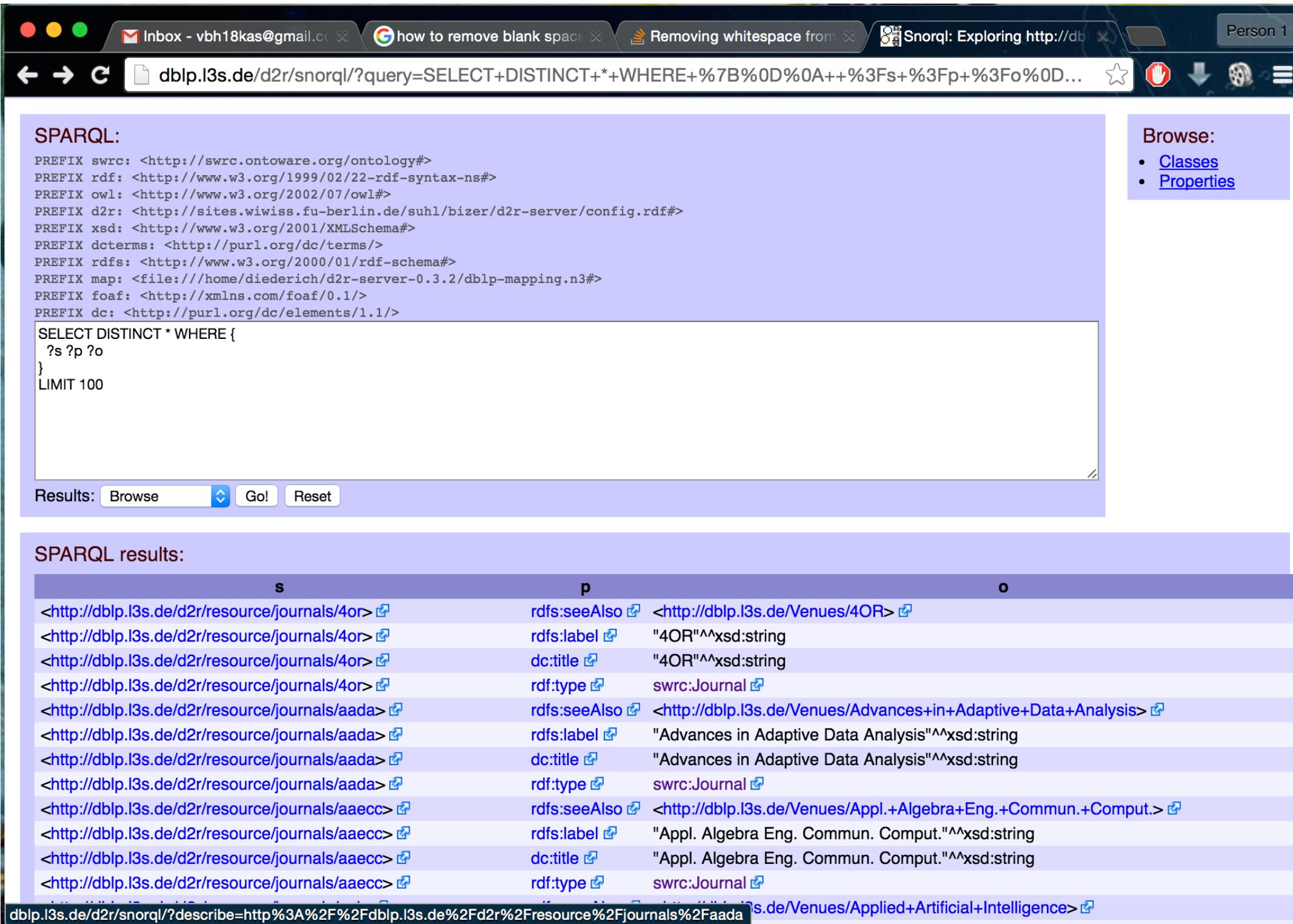
## Schema

- DBLP Dataset
  - Endpoint: <http://dblp.l3s.de/d2r/sparql>



# Schema

- DBLP Dataset



The screenshot shows a web browser window with several tabs open. The active tab is titled "Snorql: Exploring http://db..." and contains a SPARQL query editor and a results table.

**SPARQL:**

```

PREFIX swrc: <http://swrc.ontoware.org/ontology#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX d2r: <http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/config.rdf#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX map: <file:///home/diederich/d2r-server-0.3.2/dblp-mapping.n3#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>

SELECT DISTINCT * WHERE {
  ?s ?p ?o
}
LIMIT 100
  
```

**Results:** Browse   Go!   Reset

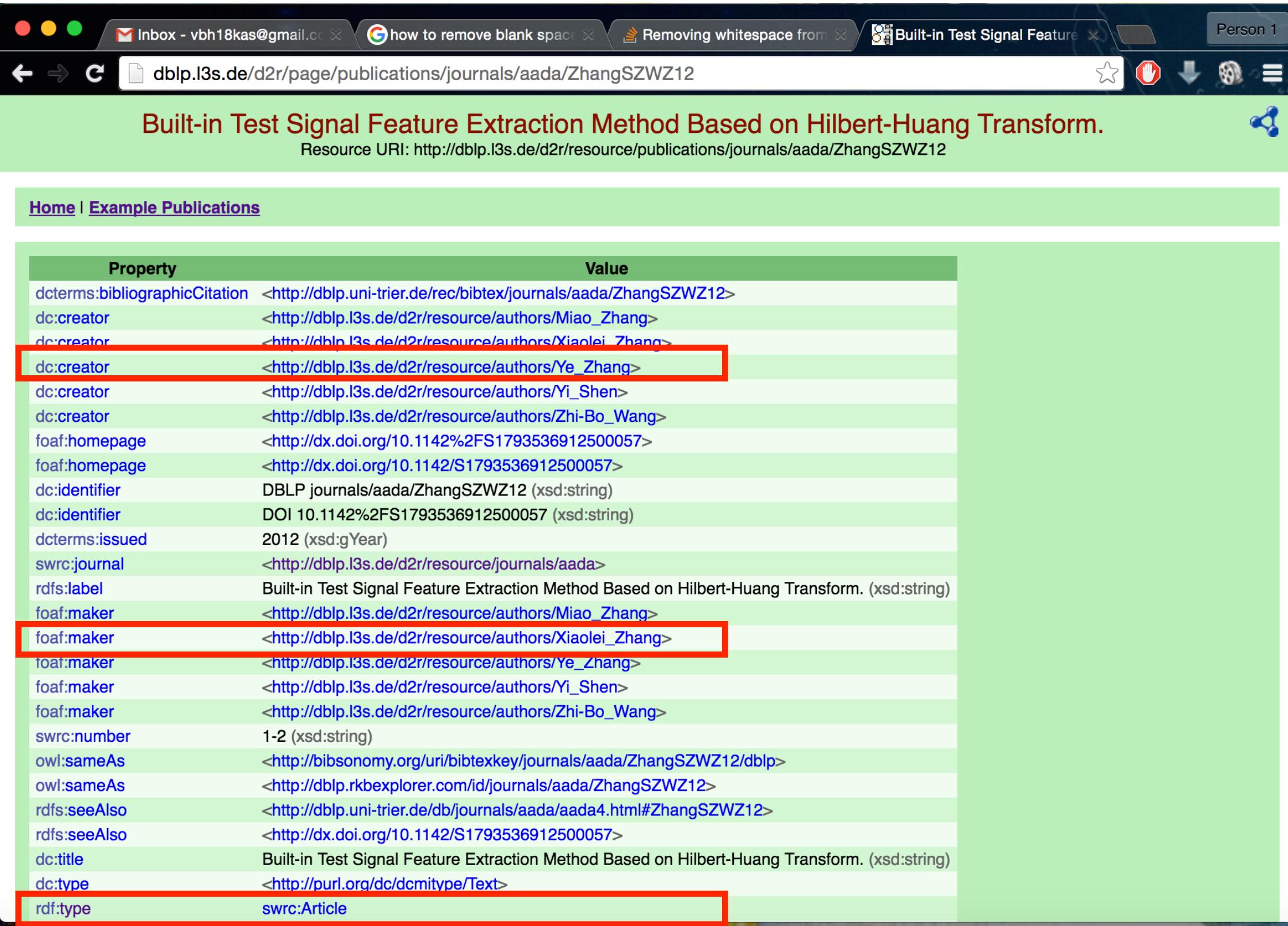
**SPARQL results:**

s	p	o
<http://dblp.l3s.de/d2r/resource/journals/4or>	rdfs:seeAlso	<http://dblp.l3s.de/Venues/4OR>
<http://dblp.l3s.de/d2r/resource/journals/4or>	rdfs:label	"4OR"^^xsd:string
<http://dblp.l3s.de/d2r/resource/journals/4or>	dc:title	"4OR"^^xsd:string
<http://dblp.l3s.de/d2r/resource/journals/4or>	rdf:type	swrc:Journal
<http://dblp.l3s.de/d2r/resource/journals/aada>	rdfs:seeAlso	<http://dblp.l3s.de/Venues/Advances+in+Adaptive+Data+Analysis>
<http://dblp.l3s.de/d2r/resource/journals/aada>	rdfs:label	"Advances in Adaptive Data Analysis"^^xsd:string
<http://dblp.l3s.de/d2r/resource/journals/aada>	dc:title	"Advances in Adaptive Data Analysis"^^xsd:string
<http://dblp.l3s.de/d2r/resource/journals/aada>	rdf:type	swrc:Journal
<http://dblp.l3s.de/d2r/resource/journals/aaecc>	rdfs:seeAlso	<http://dblp.l3s.de/Venues/Appl.+Algebra+Eng.+Commun.+Comput.>
<http://dblp.l3s.de/d2r/resource/journals/aaecc>	rdfs:label	"Appl. Algebra Eng. Commun. Comput."^^xsd:string
<http://dblp.l3s.de/d2r/resource/journals/aaecc>	dc:title	"Appl. Algebra Eng. Commun. Comput."^^xsd:string
<http://dblp.l3s.de/d2r/resource/journals/aaecc>	rdf:type	swrc:Journal

dblp.l3s.de/d2r/snorql/?describe=http%3A%2F%2Fdblp.l3s.de%2Fd2r%2Fresource%2Fjournals%2Faada

# Schema

- DBLP Dataset: (Manual) Data Analysis



Property	Value
dcterms:bibliographicCitation	<http://dblp.uni-trier.de/rec/bibtex/journals/aada/ZhangSZWZ12>
dc:creator	<http://dblp.l3s.de/d2r/resource/authors/Miao_Zhang>
dc:creator	<http://dblp.l3s.de/d2r/resource/authors/Xiaolei_Zhang>
dc:creator	<http://dblp.l3s.de/d2r/resource/authors/Ye_Zhang> <span style="border: 2px solid red; padding: 2px;"> </span>
dc:creator	<http://dblp.l3s.de/d2r/resource/authors/Yi_Shen>
dc:creator	<http://dblp.l3s.de/d2r/resource/authors/Zhi-Bo_Wang>
foaf:homepage	<http://dx.doi.org/10.1142%2FS1793536912500057>
foaf:homepage	<http://dx.doi.org/10.1142/S1793536912500057>
dc:identifier	DBLP journals/aada/ZhangSZWZ12 (xsd:string)
dc:identifier	DOI 10.1142%2FS1793536912500057 (xsd:string)
dcterms:issued	2012 (xsd:gYear)
swrc:journal	<http://dblp.l3s.de/d2r/resource/journals/aada>
rdfs:label	Built-in Test Signal Feature Extraction Method Based on Hilbert-Huang Transform. (xsd:string)
foaf:maker	<http://dblp.l3s.de/d2r/resource/authors/Miao_Zhang>
foaf:maker	<http://dblp.l3s.de/d2r/resource/authors/Xiaolei_Zhang> <span style="border: 2px solid red; padding: 2px;"> </span>
foaf:maker	<http://dblp.l3s.de/d2r/resource/authors/Ye_Zhang>
foaf:maker	<http://dblp.l3s.de/d2r/resource/authors/Yi_Shen>
foaf:maker	<http://dblp.l3s.de/d2r/resource/authors/Zhi-Bo_Wang>
swrc:number	1-2 (xsd:string)
owl:sameAs	<http://bibsonomy.org/uri/bibtexkey/journals/aada/ZhangSZWZ12/dblp>
owl:sameAs	<http://dblp.rkbexplorer.com/id/journals/aada/ZhangSZWZ12>
rdfs:seeAlso	<http://dblp.uni-trier.de/db/journals/aada/aada4.html#ZhangSZWZ12>
rdfs:seeAlso	<http://dx.doi.org/10.1142/S1793536912500057>
dc:title	Built-in Test Signal Feature Extraction Method Based on Hilbert-Huang Transform. (xsd:string)
dc:type	<http://purl.org/dc/dcmitype/Text>
rdf:type	swrc:Article <span style="border: 2px solid red; padding: 2px;"> </span>

## Querying with SPARQL



# Querying with SPARQL

- LAK Dataset : Query 1(Aggregated Query)
  - Finding Author Name, Publications, No. of Publications
  - Publications grouped together, Results arranged acc. to Count

← → ⌂ [meco.l3s.uni-hannover.de:8890/sparql](http://meco.l3s.uni-hannover.de:8890/sparql) ⭐ ⚡ ⏪ ⏴ ⏵ ⏷

Virtuoso SPARQL Query Editor [About](#) | [Namespace Prefixes](#) | [Inference rules](#)

Default Data Set Name (Graph IRI)  
  

Query Text

```

PREFIX swrc:<http://swrc.ontoware.org/ontology#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX bibo:<http://purl.org/ontology/bibo/>

SELECT ?author ?author_name count(?article) as ?num_article
    GROUP_CONCAT(?article; SEPARATOR=", ") AS ?articles
WHERE{
    ?article a bibo:Article.
    ?article foaf:maker ?author.
    ?author rdfs:label ?author_name
}GROUP BY ?author ?author_name
ORDER BY DESC(?num_article)
LIMIT 1000

```

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format:

Execution timeout:  milliseconds (values less than 1000 are ignored)

Options:  Strict checking of void variables

# Querying with SPARQL

- LAK Dataset : Query 1(Result)

Person 1			
<a href="http://meco.l3s.uni-hannover.de:8890/sparql?default-graph-uri=http%3A%2F%2Flak-conference.org%2Fpapers&amp;query=PREFIX+swrc%3A&lt;http%3A%2F%2Fdata.linkededucation.org%2Fresource%2Flak%2Fconference%2Fedm%2Fpaper%2F%3A+swrc%3Aauthor%3D%22Neil+T.+Heffernan%22+swrc%3AisPartOf%3D%22http%3A%2F%2Flak-conference.org%2Fpapers%22">http://meco.l3s.uni-hannover.de:8890/sparql?default-graph-uri=http%3A%2F%2Flak-conference.org%2Fpapers&amp;query=PREFIX+swrc%3A&lt;http%3A%2F%2Fdata.linkededucation.org%2Fresource%2Flak%2Fconference%2Fedm%2Fpaper%2F%3A+swrc%3Aauthor%3D%22Neil+T.+Heffernan%22+swrc%3AisPartOf%3D%22http%3A%2F%2Flak-conference.org%2Fpapers%22</a>			
<a href="http://data.linkededucation.org/resource/lak/person/neil-t-heffernan">http://data.linkededucation.org/resource/lak/person/neil-t-heffernan</a>	"Neil T. Heffernan"	25	<a href="http://data.linkededucation.org/resource/lak/conference/edm2009/paper/205">http://data.linkededucation.org/resource/lak/conference/edm2009/paper/205</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2008/paper/234">http://data.linkededucation.org/resource/lak/conference/edm2008/paper/234</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2010/paper/150">http://data.linkededucation.org/resource/lak/conference/edm2010/paper/150</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2013/paper/321">http://data.linkededucation.org/resource/lak/conference/edm2013/paper/321</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2010/paper/144">http://data.linkededucation.org/resource/lak/conference/edm2010/paper/144</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2011/paper/97">http://data.linkededucation.org/resource/lak/conference/edm2011/paper/97</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2011/paper/62">http://data.linkededucation.org/resource/lak/conference/edm2011/paper/62</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2012/paper/40">http://data.linkededucation.org/resource/lak/conference/edm2012/paper/40</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2010/paper/136">http://data.linkededucation.org/resource/lak/conference/edm2010/paper/136</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2012/paper/47">http://data.linkededucation.org/resource/lak/conference/edm2012/paper/47</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2010/paper/143">http://data.linkededucation.org/resource/lak/conference/edm2010/paper/143</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2013/paper/318">http://data.linkededucation.org/resource/lak/conference/edm2013/paper/318</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2008/paper/213">http://data.linkededucation.org/resource/lak/conference/edm2008/paper/213</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2011/paper/105">http://data.linkededucation.org/resource/lak/conference/edm2011/paper/105</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2008/paper/225">http://data.linkededucation.org/resource/lak/conference/edm2008/paper/225</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2009/paper/202">http://data.linkededucation.org/resource/lak/conference/edm2009/paper/202</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2010/paper/116">http://data.linkededucation.org/resource/lak/conference/edm2010/paper/116</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2014/paper/513">http://data.linkededucation.org/resource/lak/conference/edm2014/paper/513</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2014/paper/521">http://data.linkededucation.org/resource/lak/conference/edm2014/paper/521</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2014/paper/586">http://data.linkededucation.org/resource/lak/conference/edm2014/paper/586</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2014/paper/607">http://data.linkededucation.org/resource/lak/conference/edm2014/paper/607</a>
<a href="http://data.linkededucation.org/resource/lak/person/kenneth-r-koedinger">http://data.linkededucation.org/resource/lak/person/kenneth-r-koedinger</a>	"Kenneth R. Koedinger"	23	<a href="http://data.linkededucation.org/resource/lak/conference/edm2009/paper/177">http://data.linkededucation.org/resource/lak/conference/edm2009/paper/177</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2013/paper/302">http://data.linkededucation.org/resource/lak/conference/edm2013/paper/302</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2012/paper/6">http://data.linkededucation.org/resource/lak/conference/edm2012/paper/6</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2013/paper/251">http://data.linkededucation.org/resource/lak/conference/edm2013/paper/251</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2013/paper/265">http://data.linkededucation.org/resource/lak/conference/edm2013/paper/265</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2012/paper/34">http://data.linkededucation.org/resource/lak/conference/edm2012/paper/34</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2011/paper/65">http://data.linkededucation.org/resource/lak/conference/edm2011/paper/65</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2012/paper/43">http://data.linkededucation.org/resource/lak/conference/edm2012/paper/43</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2008/paper/221">http://data.linkededucation.org/resource/lak/conference/edm2008/paper/221</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2013/paper/295">http://data.linkededucation.org/resource/lak/conference/edm2013/paper/295</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2008/paper/228">http://data.linkededucation.org/resource/lak/conference/edm2008/paper/228</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2008/paper/213">http://data.linkededucation.org/resource/lak/conference/edm2008/paper/213</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2011/paper/61">http://data.linkededucation.org/resource/lak/conference/edm2011/paper/61</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2011/paper/96">http://data.linkededucation.org/resource/lak/conference/edm2011/paper/96</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2011/paper/56">http://data.linkededucation.org/resource/lak/conference/edm2011/paper/56</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2008/paper/219">http://data.linkededucation.org/resource/lak/conference/edm2008/paper/219</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2010/paper/164">http://data.linkededucation.org/resource/lak/conference/edm2010/paper/164</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2011/paper/73">http://data.linkededucation.org/resource/lak/conference/edm2011/paper/73</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2008/paper/225">http://data.linkededucation.org/resource/lak/conference/edm2008/paper/225</a>

# Querying with SPARQL

- LAK Dataset : Query 2
  - Finding Author Name and Organization

Virtuoso SPARQL Query Editor

About | Namespace Prefixes | Inference rules

Default Data Set Name (Graph IRI)

Query Text

```

PREFIX swrc:<http://swrc.ontoware.org/ontology#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX bibo:<http://purl.org/ontology/bibo/>

SELECT ?author2 ?author_name2 ?organization ?organization_name
WHERE{
    ?author2 swrc:affiliation ?organization.
    ?author2 rdfs:label ?author_name2.
    ?author2 rdfs:label ?organization_name.
}

```

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format:

Execution timeout:  milliseconds (values less than 1000 are ignored)

Options:  Strict checking of void variables

(The result can only be sent back to browser, not saved on the server, see [details](#))

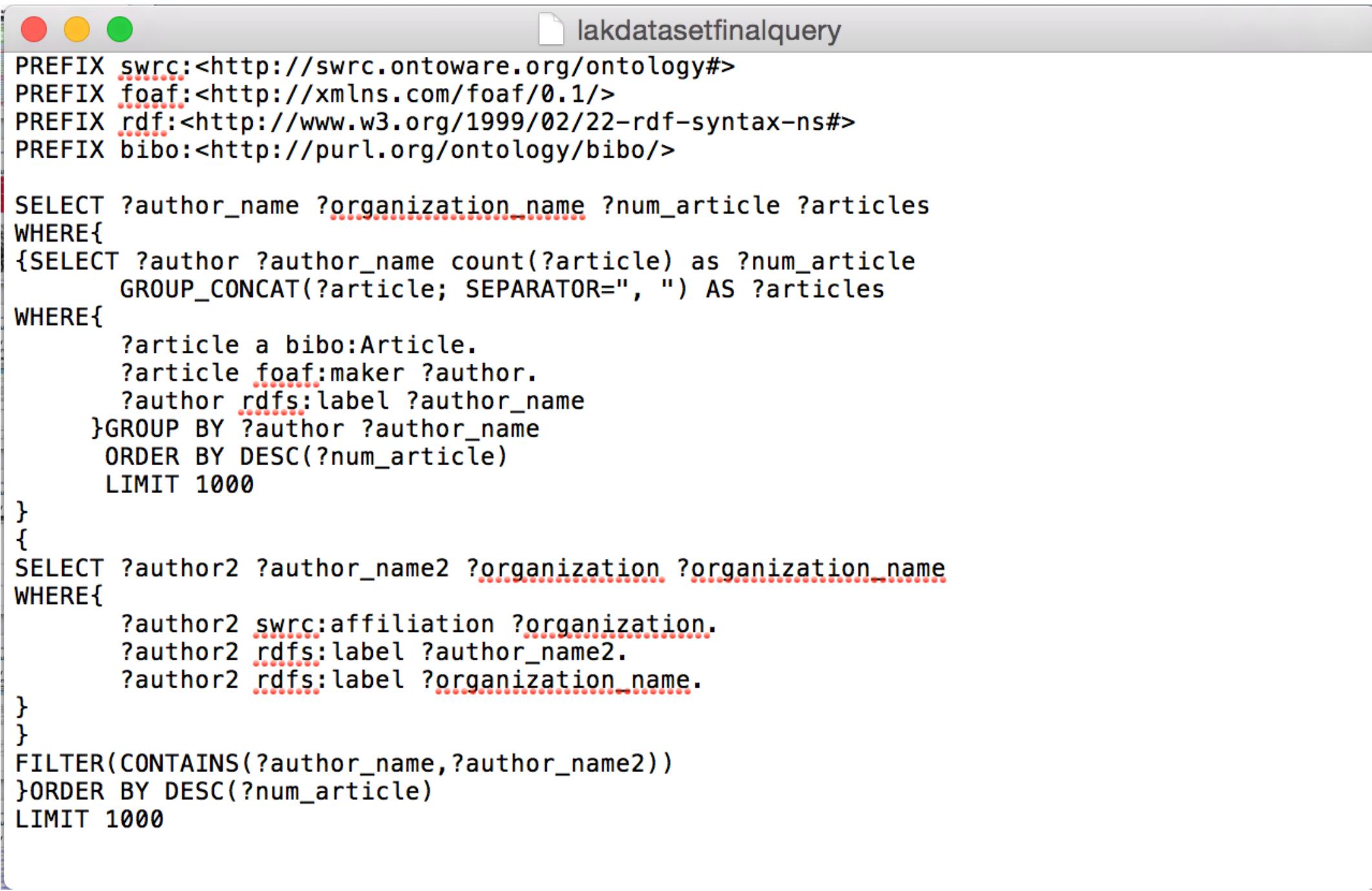
# Querying with SPARQL

- LAK Dataset : Query 2(Result)

"Behnam Taraghi"	<a href="http://data.linkededucation.org/resource/lak/organization/graz-university-of-technology">http://data.linkededucation.org/resource/lak/organization/graz-university-of-technology</a>
"Belinda Chiera"	<a href="http://data.linkededucation.org/resource/lak/organization/university-of-south-australia">http://data.linkededucation.org/resource/lak/organization/university-of-south-australia</a>
"Bob Wadholm"	<a href="http://data.linkededucation.org/resource/lak/organization/university-of-missouri">http://data.linkededucation.org/resource/lak/organization/university-of-missouri</a>
"Bodong Chen"	<a href="http://data.linkededucation.org/resource/lak/organization/oise/university-of-toronto">http://data.linkededucation.org/resource/lak/organization/oise/university-of-toronto</a>
"Brian Sulcer"	<a href="http://data.linkededucation.org/resource/lak/organization/vanderbilt-university">http://data.linkededucation.org/resource/lak/organization/vanderbilt-university</a>
"Carleton Coffrin"	<a href="http://data.linkededucation.org/resource/lak/organization/national-ict-australia">http://data.linkededucation.org/resource/lak/organization/national-ict-australia</a>
"Caroline Haythornthwaite"	<a href="http://data.linkededucation.org/resource/lak/organization/university-of-british-columbia">http://data.linkededucation.org/resource/lak/organization/university-of-british-columbia</a>
"Cassandra Colvin"	<a href="http://data.linkededucation.org/resource/lak/organization/university-of-south-australia">http://data.linkededucation.org/resource/lak/organization/university-of-south-australia</a>
"Chin-hwa Kuo"	<a href="http://data.linkededucation.org/resource/lak/organization/tamkang-university">http://data.linkededucation.org/resource/lak/organization/tamkang-university</a>

# Querying with SPARQL

- LAK Dataset : Query 3
  - Combining Results from Query 1 and Query 2
  - Filtering Results based on Author (Sorted by Num\_Publications)



```

PREFIX swrc:<http://swrc.ontoware.org/ontology#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX bibo:<http://purl.org/ontology/bibo/>

SELECT ?author_name ?organization_name ?num_article ?articles
WHERE{
{SELECT ?author ?author_name count(?article) as ?num_article
GROUP_CONCAT(?article; SEPARATOR=", ") AS ?articles
WHERE{
?article a bibo:Article.
?article foaf:maker ?author.
?author rdfs:label ?author_name
}GROUP BY ?author ?author_name
ORDER BY DESC(?num_article)
LIMIT 1000
}
{
SELECT ?author2 ?author_name2 ?organization ?organization_name
WHERE{
?author2 swrc:affiliation ?organization.
?author2 rdfs:label ?author_name2.
?author2 rdfs:label ?organization_name.
}
}
FILTER(CONTAINS(?author_name,?author_name2))
}ORDER BY DESC(?num_article)
LIMIT 1000

```

# Querying with SPARQL

- LAK Dataset : Query 3(Result)

"Zachary A. Pardos"	<a href="http://data.linkededucation.org/resource/lak/organization/uc-berkeley">http://data.linkededucation.org/resource/lak/organization/uc-berkeley</a>	14	<a href="http://data.linkededucation.org/resource/lak/conference/edm2014/paper/60">http://data.linkededucation.org/resource/lak/conference/edm2014/paper/60</a> , <a href="http://data.linkededucation.org/resource/lak/journal/jla201401/paper/479">http://data.linkededucation.org/resource/lak/journal/jla201401/paper/479</a> , <a href="http://data.linkededucation.org/resource/lak/journal/jla201401/paper/479">http://data.linkededucation.org/resource/lak/journal/jla201401/paper/479</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2013/paper/27">http://data.linkededucation.org/resource/lak/conference/edm2013/paper/27</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2009/paper/19">http://data.linkededucation.org/resource/lak/conference/edm2009/paper/19</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2008/paper/23">http://data.linkededucation.org/resource/lak/conference/edm2008/paper/23</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2010/paper/15">http://data.linkededucation.org/resource/lak/conference/edm2010/paper/15</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2011/paper/97">http://data.linkededucation.org/resource/lak/conference/edm2011/paper/97</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2011/paper/62">http://data.linkededucation.org/resource/lak/conference/edm2011/paper/62</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2012/paper/47">http://data.linkededucation.org/resource/lak/conference/edm2012/paper/47</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2012/paper/2">http://data.linkededucation.org/resource/lak/conference/edm2012/paper/2</a> , <a href="http://data.linkededucation.org/resource/lak/conference/lak2013/paper/106">http://data.linkededucation.org/resource/lak/conference/lak2013/paper/106</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2013/paper/33">http://data.linkededucation.org/resource/lak/conference/edm2013/paper/33</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2012/paper/19">http://data.linkededucation.org/resource/lak/conference/edm2012/paper/19</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2014/paper/60">http://data.linkededucation.org/resource/lak/conference/edm2014/paper/60</a>
"Danielle S. Mcnamara"	<a href="http://data.linkededucation.org/resource/lak/organization/arizona-state-university">http://data.linkededucation.org/resource/lak/organization/arizona-state-university</a>	12	<a href="http://data.linkededucation.org/resource/lak/conference/edm2013/paper/27">http://data.linkededucation.org/resource/lak/conference/edm2013/paper/27</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2013/paper/27">http://data.linkededucation.org/resource/lak/conference/edm2013/paper/27</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2013/paper/28">http://data.linkededucation.org/resource/lak/conference/edm2013/paper/28</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2012/paper/38">http://data.linkededucation.org/resource/lak/conference/edm2012/paper/38</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2013/paper/29">http://data.linkededucation.org/resource/lak/conference/edm2013/paper/29</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2013/paper/30">http://data.linkededucation.org/resource/lak/conference/edm2013/paper/30</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2014/paper/50">http://data.linkededucation.org/resource/lak/conference/edm2014/paper/50</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2014/paper/54">http://data.linkededucation.org/resource/lak/conference/edm2014/paper/54</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2014/paper/55">http://data.linkededucation.org/resource/lak/conference/edm2014/paper/55</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2014/paper/56">http://data.linkededucation.org/resource/lak/conference/edm2014/paper/56</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2014/paper/57">http://data.linkededucation.org/resource/lak/conference/edm2014/paper/57</a> , <a href="http://data.linkededucation.org/resource/lak/conference/edm2014/paper/58">http://data.linkededucation.org/resource/lak/conference/edm2014/paper/58</a>
"Shane Dawson"	<a href="http://data.linkededucation.org/resource/lak/organization/university-of-south-australia">http://data.linkededucation.org/resource/lak/organization/university-of-south-australia</a>	11	<a href="http://data.linkededucation.org/resource/lak/journal/jla201401/paper/480">http://data.linkededucation.org/resource/lak/journal/jla201401/paper/480</a> , <a href="http://data.linkededucation.org/resource/lak/journal/jla201401/paper/480">http://data.linkededucation.org/resource/lak/journal/jla201401/paper/480</a> , <a href="http://data.linkededucation.org/resource/lak/journal/jla201402/paper/485">http://data.linkededucation.org/resource/lak/journal/jla201402/paper/485</a> , <a href="http://data.linkededucation.org/resource/lak/conference/lak2014/paper/636">http://data.linkededucation.org/resource/lak/conference/lak2014/paper/636</a> , <a href="http://data.linkededucation.org/resource/lak/conference/lak2014/paper/649">http://data.linkededucation.org/resource/lak/conference/lak2014/paper/649</a> , <a href="http://data.linkededucation.org/resource/lak/conference/lak2014/paper/650">http://data.linkededucation.org/resource/lak/conference/lak2014/paper/650</a> , <a href="http://data.linkededucation.org/resource/lak/specialissue/jets12/paper/69">http://data.linkededucation.org/resource/lak/specialissue/jets12/paper/69</a> , <a href="http://data.linkededucation.org/resource/lak/conference/lak2011/paper/62">http://data.linkededucation.org/resource/lak/conference/lak2011/paper/62</a> , <a href="http://data.linkededucation.org/resource/lak/conference/lak2013/paper/111">http://data.linkededucation.org/resource/lak/conference/lak2013/paper/111</a> , <a href="http://data.linkededucation.org/resource/lak/conference/lak2011/paper/51">http://data.linkededucation.org/resource/lak/conference/lak2011/paper/51</a> , <a href="http://data.linkededucation.org/resource/lak/journal/jla201401/paper/480">http://data.linkededucation.org/resource/lak/journal/jla201401/paper/480</a>

# Querying with SPARQL

# Querying with SPARQL

## Query 1



# Querying with SPARQL

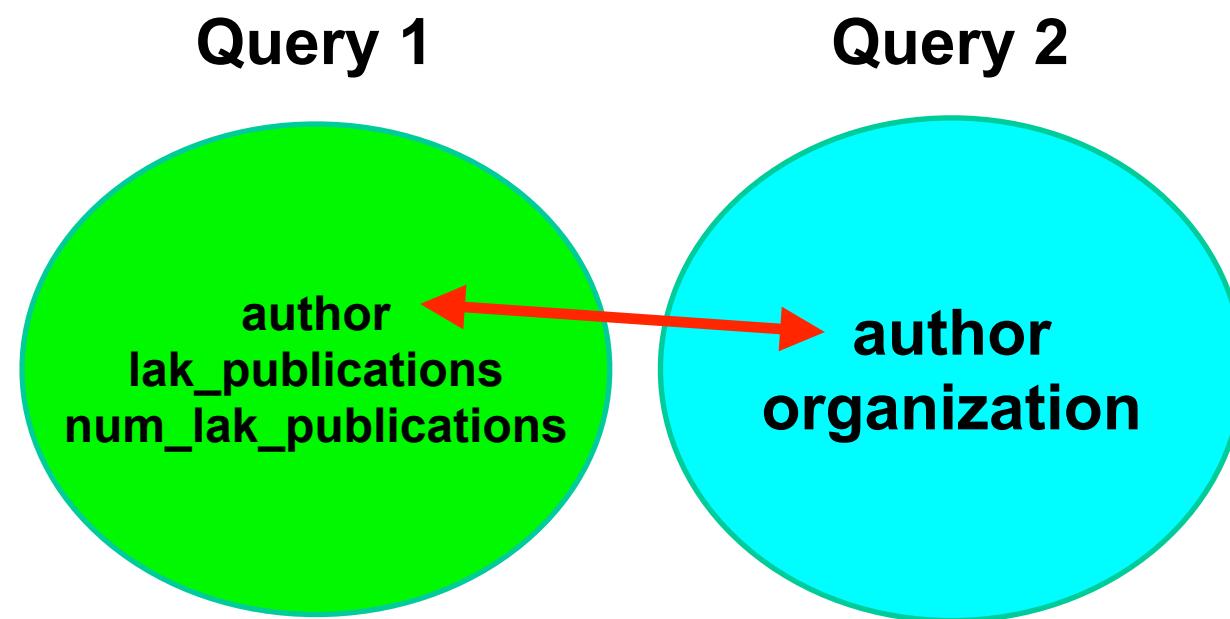
**Query 1**



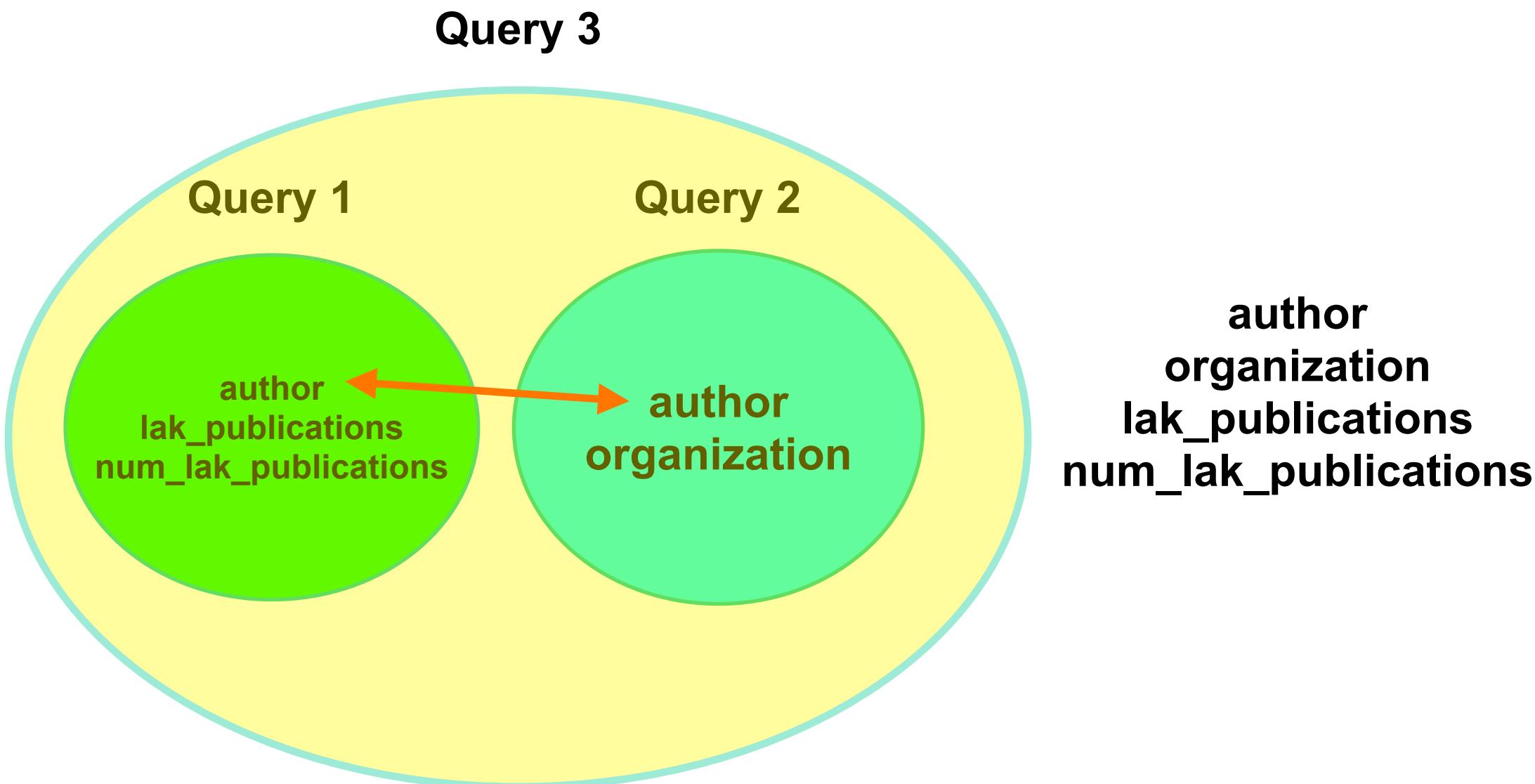
**Query 2**



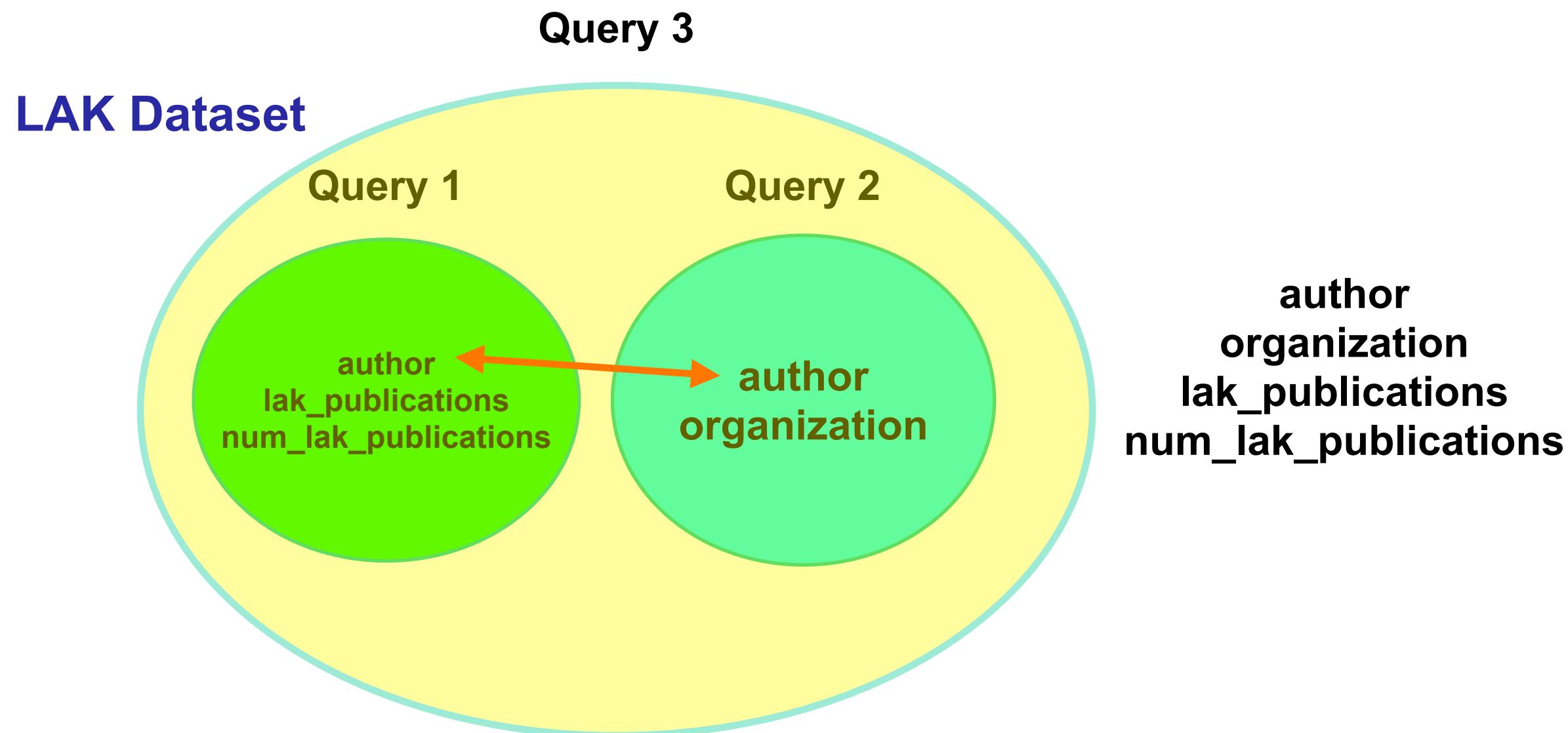
# Querying with SPARQL



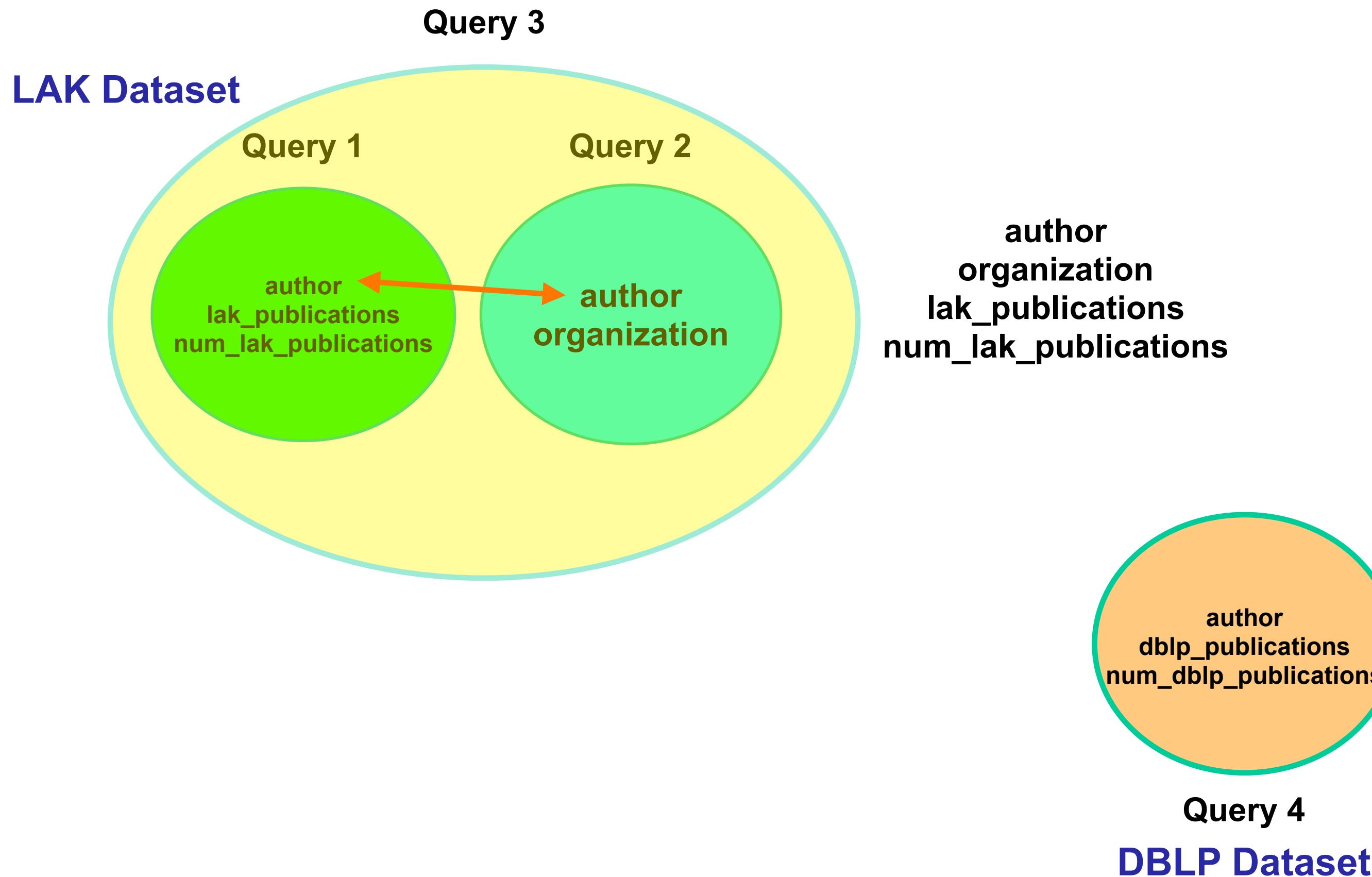
# Querying with SPARQL



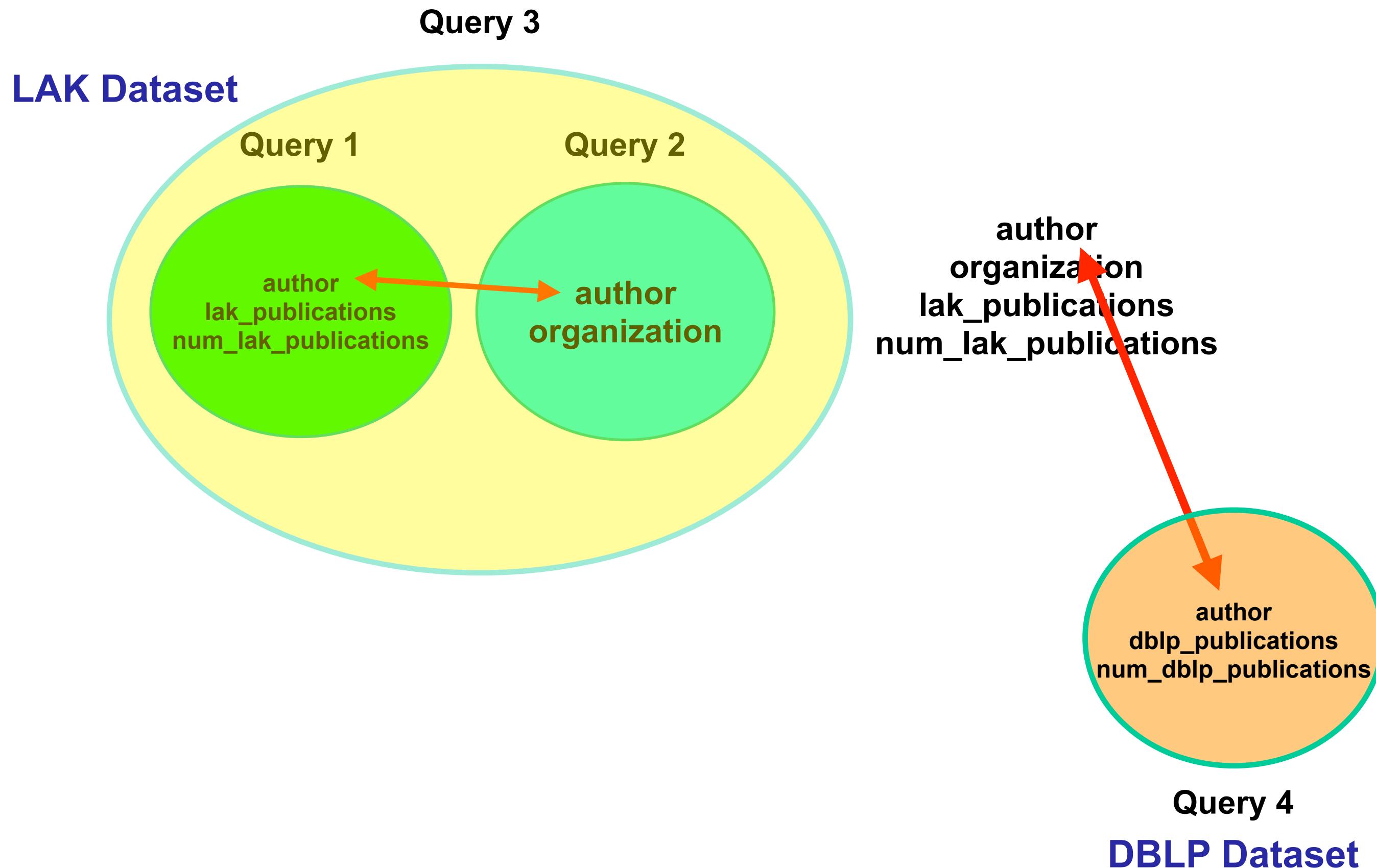
# Querying with SPARQL



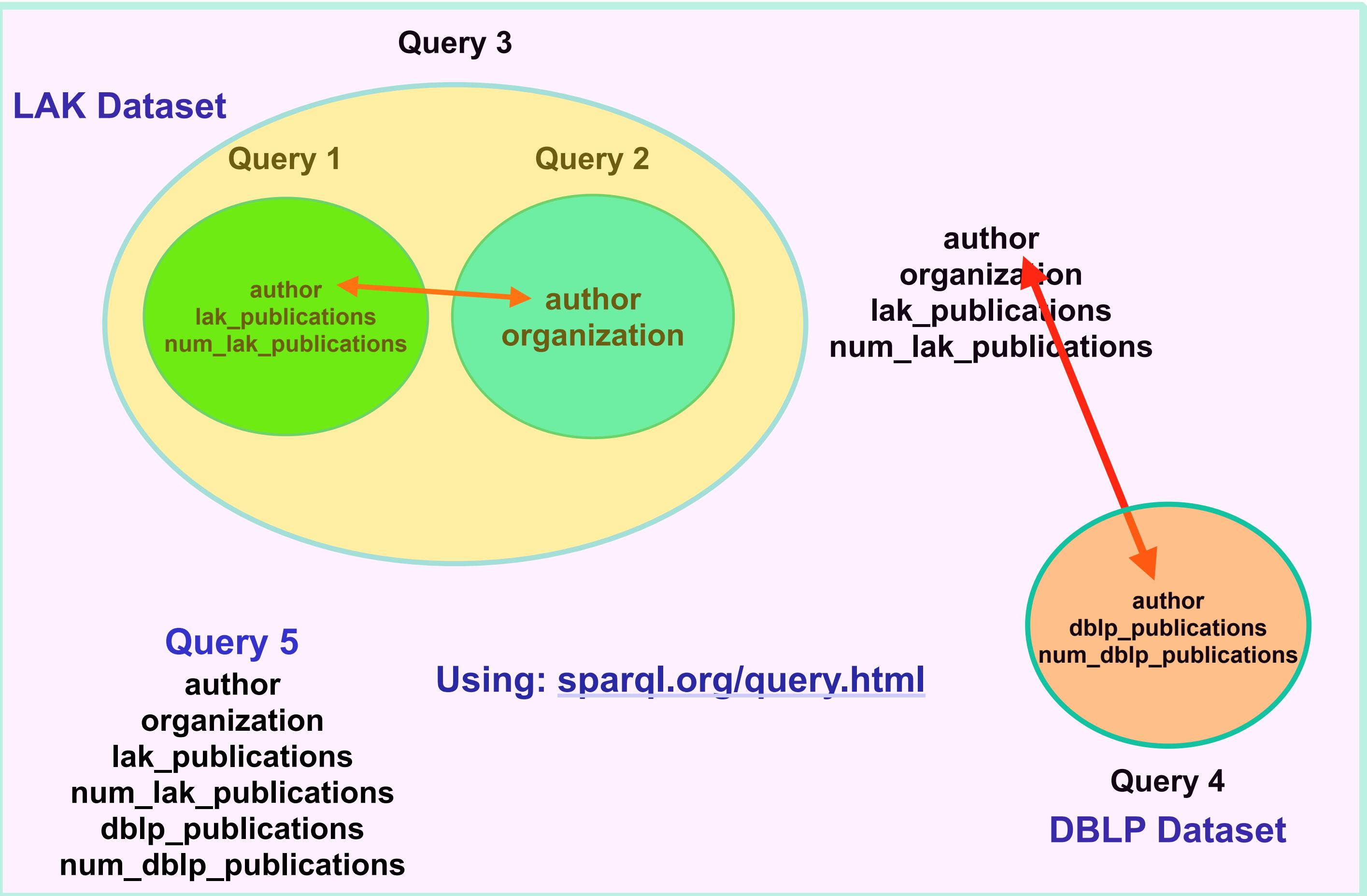
# Querying with SPARQL



# Querying with SPARQL

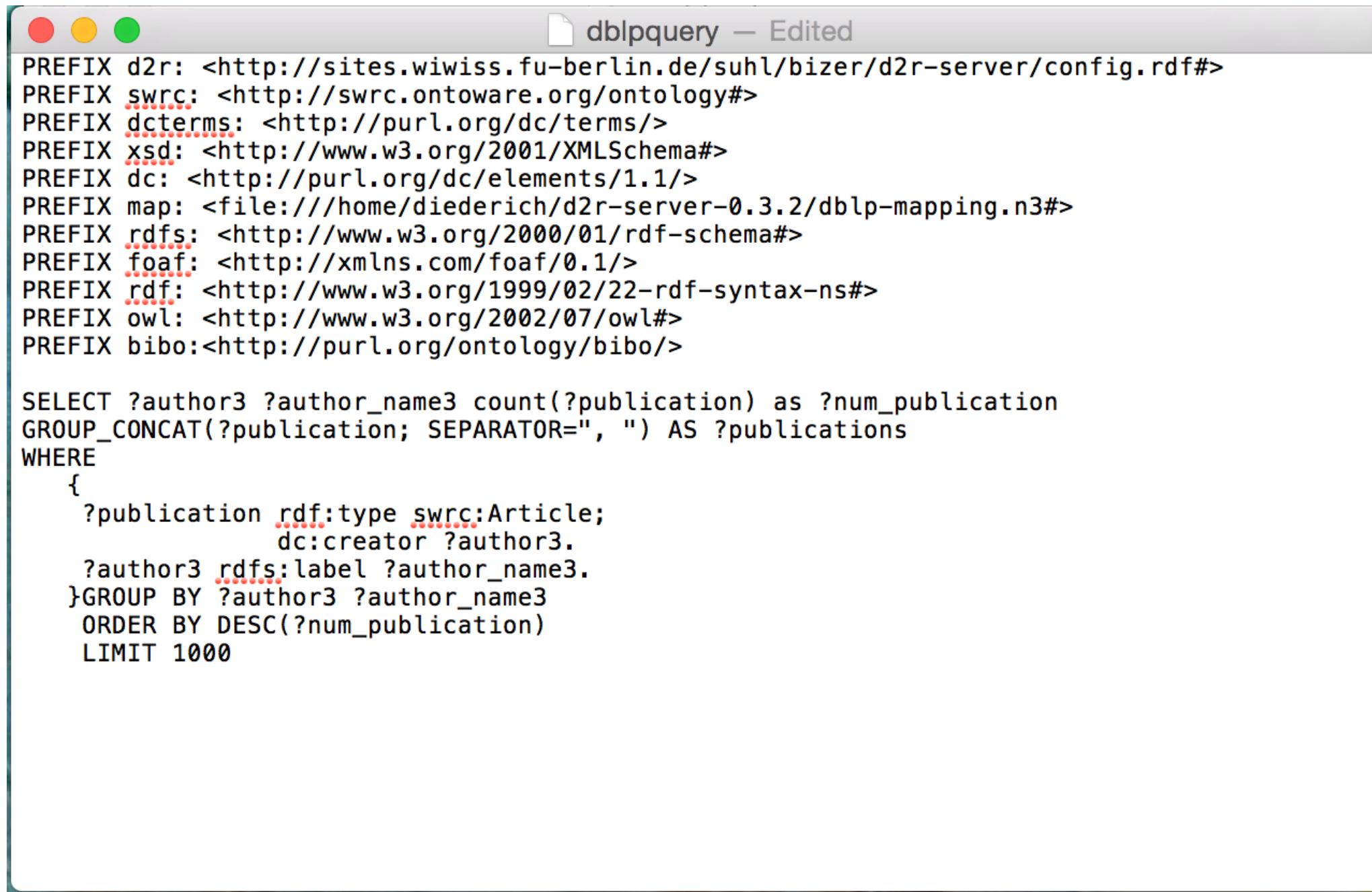


# Querying with SPARQL



# Querying with SPARQL

- DBLP Dataset : Query 4(Aggregated Query)
  - Finding Author Name, Publications, No. of Publications
  - Publications grouped together, Results arranged acc. to Count



The screenshot shows a SPARQL query editor window titled "dblpquery – Edited". The query itself is as follows:

```

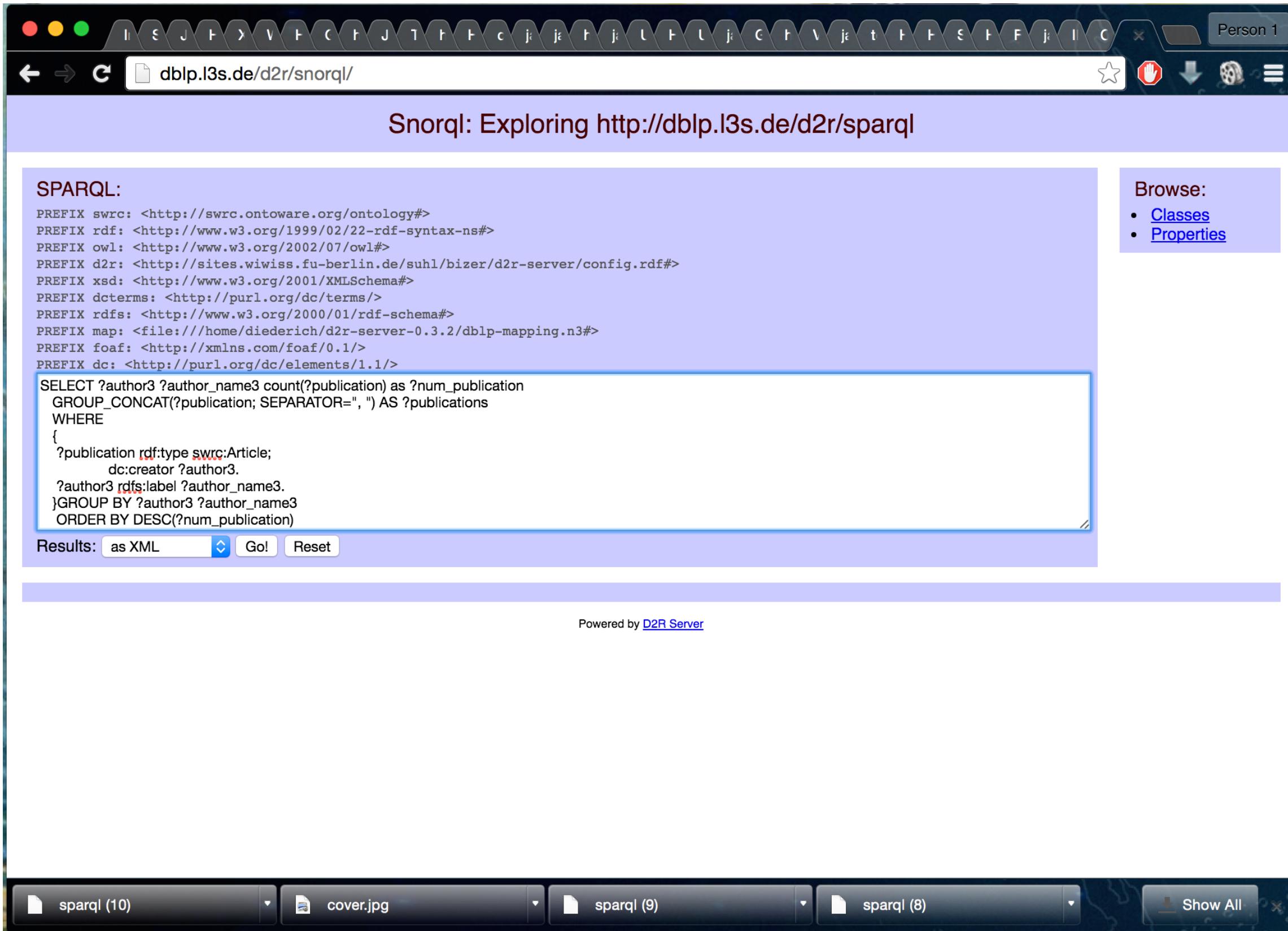
PREFIX d2r: <http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/config.rdf#>
PREFIX swrc: <http://swrc.ontoware.org/ontology#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX map: <file:///home/diederich/d2r-server-0.3.2/dblp-mapping.n3#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX bibo:<http://purl.org/ontology/bibo/>

SELECT ?author3 ?author_name3 count(?publication) as ?num_publication
GROUP_CONCAT(?publication; SEPARATOR=", ") AS ?publications
WHERE
{
    ?publication rdf:type swrc:Article;
                  dc:creator ?author3.
    ?author3 rdfs:label ?author_name3.
} GROUP BY ?author3 ?author_name3
ORDER BY DESC(?num_publication)
LIMIT 1000

```

# Querying with SPARQL

- DBLP Dataset : Query 4



The screenshot shows a web browser window with the following details:

- Address Bar:** dblp.l3s.de/d2r/snорql/
- Title Bar:** Snорql: Exploring http://dblp.l3s.de/d2r/sparql
- Left Panel (SPARQL):**
  - Prefix declarations (swrc, rdf, owl, d2r, xsd, dcterms, rdfs, map, foaf, dc).
  - SPARQL query:

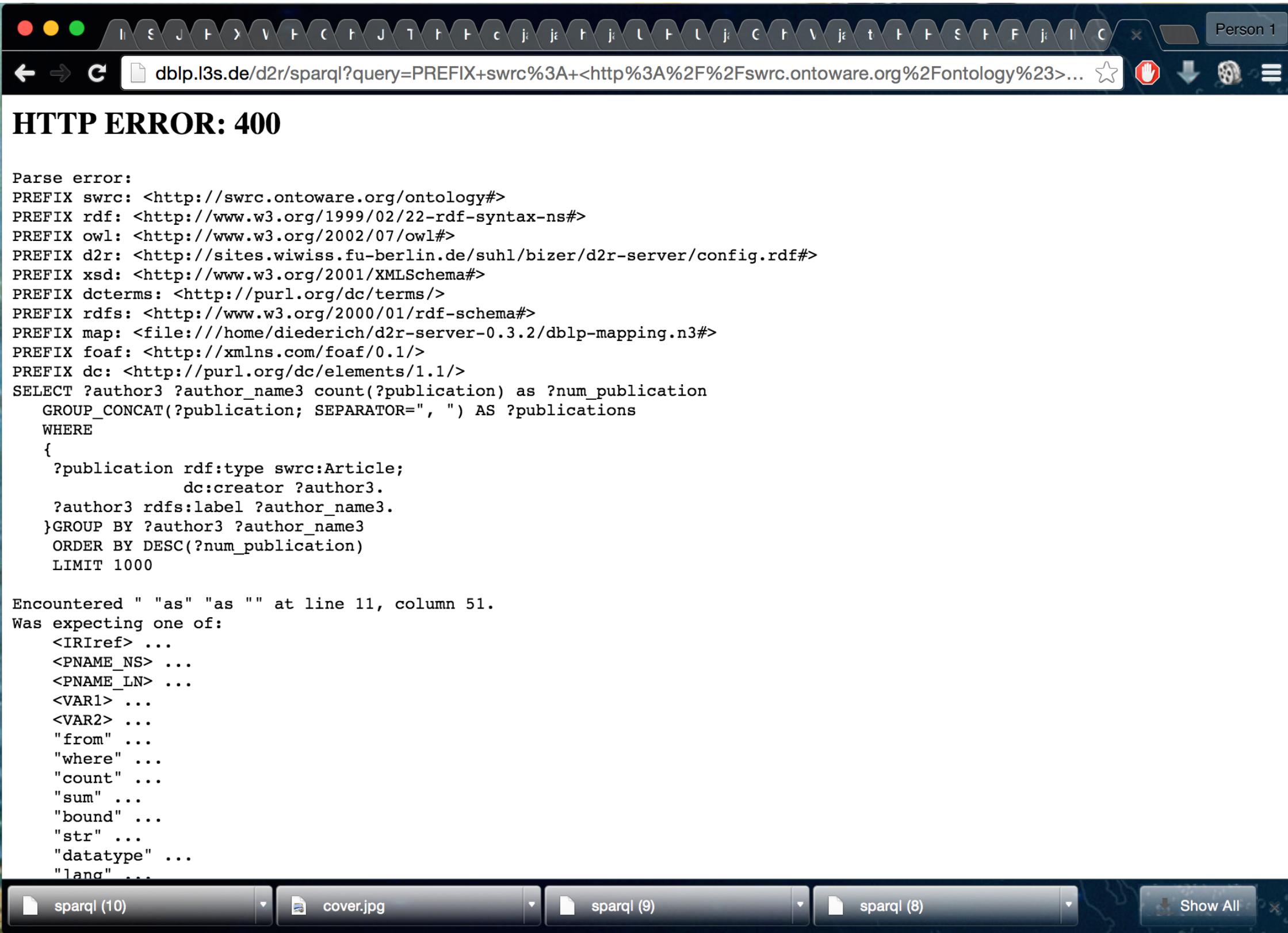
```

SELECT ?author3 ?author_name3 count(?publication) as ?num_publication
GROUP_CONCAT(?publication; SEPARATOR=", ") AS ?publications
WHERE
{
  ?publication rdf:type swrc:Article;
    dc:creator ?author3.
  ?author3 rdfs:label ?author_name3.
}GROUP BY ?author3 ?author_name3
ORDER BY DESC(?num_publication)

```
  - Results dropdown: as XML, Go!, Reset.
- Right Panel (Browse):**
  - Person 1
  - Browse: Classes, Properties
- Bottom Status Bar:** Powered by D2R Server
- Bottom Taskbar:** sparql (10), cover.jpg, sparql (9), sparql (8), Show All

# Querying with SPARQL

- DBLP Dataset : Query 4(Result)



The screenshot shows a web browser window with a dark theme. The address bar contains the URL: `dblp.l3s.de/d2r/sparql?query=PREFIX+swrc%3A+<http%3A%2F%2Fswrc.ontoware.org%2Fontology%23>...`. The main content area displays an **HTTP ERROR: 400** page. The error message is as follows:

```

Parse error:
PREFIX swrc: <http://swrc.ontoware.org/ontology#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX d2r: <http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/config.rdf#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX map: <file:///home/diederich/d2r-server-0.3.2/dblp-mapping.n3#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
SELECT ?author3 ?author_name3 count(?publication) as ?num_publication
  GROUP_CONCAT(?publication; SEPARATOR=", ") AS ?publications
  WHERE
  {
    ?publication rdf:type swrc:Article;
      dc:creator ?author3.
    ?author3 rdfs:label ?author_name3.
  }GROUP BY ?author3 ?author_name3
  ORDER BY DESC(?num_publication)
  LIMIT 1000

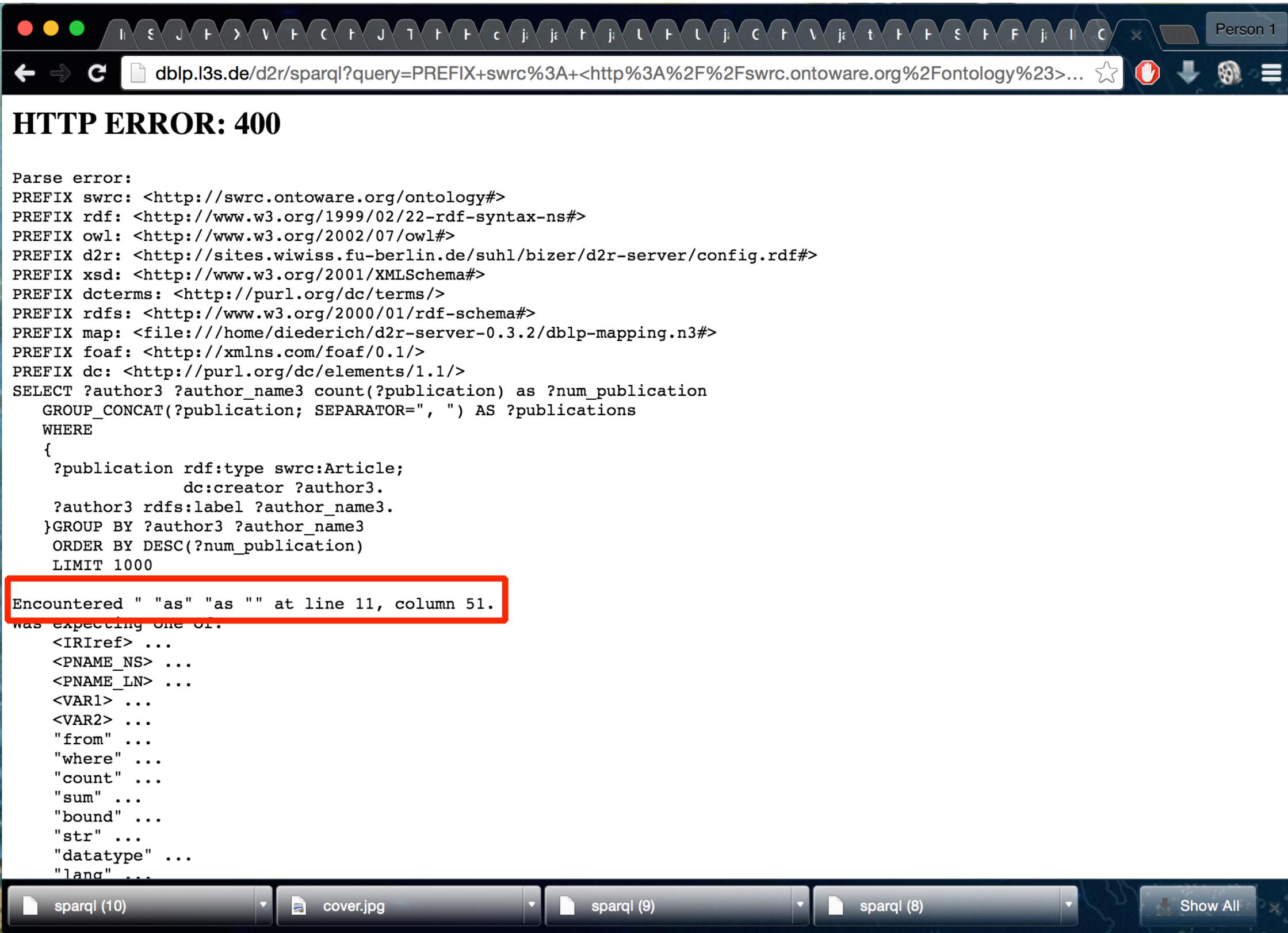
Encountered " "as" "as " " at line 11, column 51.
Was expecting one of:
<IRIref> ...
<PNAME_NS> ...
<PNAME_LN> ...
<VAR1> ...
<VAR2> ...
"from" ...
"where" ...
"count" ...
"sum" ...
"bound" ...
"str" ...
"datatype" ...
"lang" ...

```

The bottom of the browser window shows several tabs: `sparql (10)`, `cover.jpg`, `sparql (9)`, `sparql (8)`, and `Show All`.

# Querying with SPARQL

- DBLP Dataset : Query 4(Result)



The screenshot shows a web browser window with a dark theme. The address bar contains the URL: `dblp.l3s.de/d2r/sparql?query=PREFIX+swrc%3A+<http%3A%2F%2Fswrc.ontoware.org%2Fontology%23>...`. The main content area displays an **HTTP ERROR: 400** message. Below it, a **Parse error:** message is shown, followed by the SPARQL query text. A red box highlights the error message: **Encountered " "as" "as " " at line 11, column 51.** The query text includes various prefixes and a complex SELECT clause. At the bottom, there are several tabs: `sparql (10)`, `cover.jpg`, `sparql (9)`, `sparql (8)`, and `Show All`.

```

Parse error:
PREFIX swrc: <http://swrc.ontoware.org/ontology#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX d2r: <http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/config.rdf#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX map: <file:///home/diederich/d2r-server-0.3.2/dblp-mapping.n3#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
SELECT ?author3 ?author_name3 count(?publication) as ?num_publication
  GROUP_CONCAT(?publication; SEPARATOR=", ") AS ?publications
  WHERE
  {
    ?publication rdf:type swrc:Article;
      dc:creator ?author3.
    ?author3 rdfs:label ?author_name3
  }GROUP BY ?author3 ?author_name3
  ORDER BY DESC(?num_publication)
  LIMIT 1000
Encountered " "as" "as " " at line 11, column 51.
was expecting one of:
<IRIref> ...
<PNAME_NS> ...
<PNAME_LN> ...
<VAR1> ...
<VAR2> ...
"from" ...
"where" ...
"count" ...
"sum" ...
"bound" ...
"str" ...
"datatype" ...
"lang" ...

```

# Querying with SPARQL

- DBLP Dataset : Query for Author and Publication (Query A)

Snorql: Exploring <http://dblp.l3s.de/d2r/sparql>

**SPARQL:**

```
PREFIX d2r: <http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/config.rdf#>
PREFIX swrc: <http://swrc.ontoware.org/ontology#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX map: <file:///home/diederich/d2r-server-0.3.2/dblp-mapping.n3#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
```

```
SELECT ?author ?publication
WHERE {
    ?publication rdf:type swrc:Article;
                 dc:creator ?author.
    ?author rdfs:label ?author_name
}
```

**Browse:**

- [Classes](#)
- [Properties](#)

Results:

**SPARQL results:**

author	publication
<a href="http://dblp.l3s.de/d2r/resource/authors/Michael_Ley">http://dblp.l3s.de/d2r/resource/authors/Michael_Ley</a>	<a href="http://dblp.l3s.de/d2r/resource/publications/persons/Ley95">http://dblp.l3s.de/d2r/resource/publications/persons/Ley95</a>
<a href="http://dblp.l3s.de/d2r/resource/authors/Rita_Ley">http://dblp.l3s.de/d2r/resource/authors/Rita_Ley</a>	<a href="http://dblp.l3s.de/d2r/resource/publications/persons/LeyCHM11">http://dblp.l3s.de/d2r/resource/publications/persons/LeyCHM11</a>
<a href="http://dblp.l3s.de/d2r/resource/authors/Markus_Casper">http://dblp.l3s.de/d2r/resource/authors/Markus_Casper</a>	<a href="http://dblp.l3s.de/d2r/resource/publications/persons/LeyCHM11">http://dblp.l3s.de/d2r/resource/publications/persons/LeyCHM11</a>
<a href="http://dblp.l3s.de/d2r/resource/authors/Hugo_Hellebrand">http://dblp.l3s.de/d2r/resource/authors/Hugo_Hellebrand</a>	<a href="http://dblp.l3s.de/d2r/resource/publications/persons/LeyCHM11">http://dblp.l3s.de/d2r/resource/publications/persons/LeyCHM11</a>
<a href="http://dblp.l3s.de/d2r/resource/authors/Ralf_Merz">http://dblp.l3s.de/d2r/resource/authors/Ralf_Merz</a>	<a href="http://dblp.l3s.de/d2r/resource/publications/persons/LeyCHM11">http://dblp.l3s.de/d2r/resource/publications/persons/LeyCHM11</a>
<a href="http://dblp.l3s.de/d2r/resource/authors/Markus_Casper">http://dblp.l3s.de/d2r/resource/authors/Markus_Casper</a>	<a href="http://dblp.l3s.de/d2r/resource/publications/persons/CasperGGGHLR12">http://dblp.l3s.de/d2r/resource/publications/persons/CasperGGGHLR12</a>
<a href="http://dblp.l3s.de/d2r/resource/authors/Gayane_Grigoryan">http://dblp.l3s.de/d2r/resource/authors/Gayane_Grigoryan</a>	<a href="http://dblp.l3s.de/d2r/resource/publications/persons/CasperGGGHLR12">http://dblp.l3s.de/d2r/resource/publications/persons/CasperGGGHLR12</a>
<a href="http://dblp.l3s.de/d2r/resource/authors/Oliver_Gronz">http://dblp.l3s.de/d2r/resource/authors/Oliver_Gronz</a>	<a href="http://dblp.l3s.de/d2r/resource/publications/persons/CasperGGGHLR12">http://dblp.l3s.de/d2r/resource/publications/persons/CasperGGGHLR12</a>
<a href="http://dblp.l3s.de/d2r/resource/authors/Oliver_Gutjahr">http://dblp.l3s.de/d2r/resource/authors/Oliver_Gutjahr</a>	<a href="http://dblp.l3s.de/d2r/resource/publications/persons/CasperGGGHLR12">http://dblp.l3s.de/d2r/resource/publications/persons/CasperGGGHLR12</a>
<a href="http://dblp.l3s.de/d2r/resource/authors/G%C3%BCnther_Heinemann">http://dblp.l3s.de/d2r/resource/authors/G%C3%BCnther_Heinemann</a>	<a href="http://dblp.l3s.de/d2r/resource/publications/persons/CasperGGGHLR12">http://dblp.l3s.de/d2r/resource/publications/persons/CasperGGGHLR12</a>
<a href="http://dblp.l3s.de/d2r/resource/authors/Rita_Ley">http://dblp.l3s.de/d2r/resource/authors/Rita_Ley</a>	<a href="http://dblp.l3s.de/d2r/resource/publications/persons/CasperGGGHLR12">http://dblp.l3s.de/d2r/resource/publications/persons/CasperGGGHLR12</a>

# Querying with SPARQL

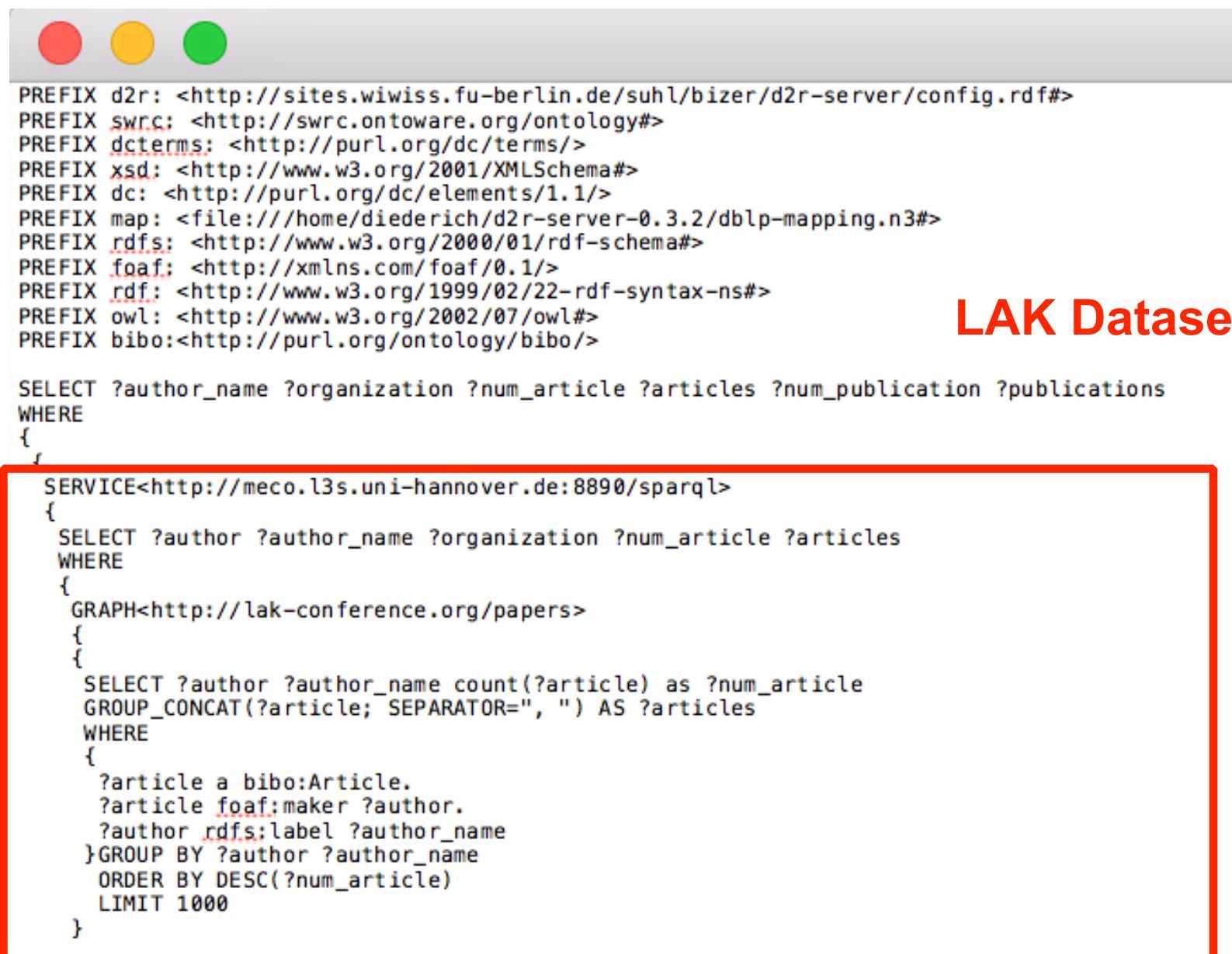
- Both Datasets : Query 5(Federated Query)
    - Finding Author Name, Publications, Num\_Publications (DBLP,LAK)
    - Results arranged acc. to Count (No. of LAK Publications)

```
PREFIX d2r: <http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/config.rdf#>
PREFIX swrc: <http://swrc.ontoware.org/ontology#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX map: <file:///home/diederich/d2r-server-0.3.2/dblp-mapping.n3#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX bibo:<http://purl.org/ontology/bibo/>

SELECT ?author_name ?organization ?num_article ?articles ?num_publication ?publications
WHERE
{
{
  SERVICE<http://meco.l3s.uni-hannover.de:8890/sparql>
  {
    SELECT ?author ?author_name ?organization ?num_article ?articles
    WHERE
    {
      GRAPH<http://lak-conference.org/papers>
      {
        SELECT ?author ?author_name count(?article) as ?num_article
        GROUP_CONCAT(?article; SEPARATOR=", ") AS ?articles
        WHERE
        {
          ?article a bibo:Article.
          ?article foaf:maker ?author.
          ?author rdfs:label ?author_name
        }GROUP BY ?author ?author_name
        ORDER BY DESC(?num_article)
        LIMIT 1000
      }
    }
  }
}
  {
    SELECT ?author2 ?author_name2 ?organization
    WHERE
    {
      ?author2 swrc:affiliation ?organization.
      ?author2 rdfs:label ?author_name2
    }
  }
}
  FILTER(CONTAINS(?author_name,?author_name2))
}
}
  {
    SERVICE<http://dblp.l3s.de/d2r/sparql>
    {
      SELECT ?author3 ?author_name3 count(?publication) as ?num_publication
      GROUP_CONCAT(?publication; SEPARATOR=", ") AS ?publications
      WHERE
      {
        ?publication rdf:type swrc:Article;
                      dc:creator ?author3.
        ?author3 rdfs:label ?author_name3.
      }GROUP BY ?author3 ?author_name3
      ORDER BY DESC(?num_publication)
      LIMIT 1000
    }
}
  FILTER(CONTAINS(?author_name,?author_name3))
}
```

# Querying with SPARQL

- Both Datasets : Query 5(Federated Query)
    - Finding Author Name, Publications, Num\_Publications (DBLP,LAK)
    - Results arranged acc. to Count (No. of LAK Publications)



## I AK Dataset

```
{  
  SELECT ?author2 ?author_name2 ?organization  
  WHERE  
  {  
    ?author2 swrc:affiliation ?organization.  
    ?author2 rdfs:label ?author_name2  
  }  
  }  
  }  
  FILTER(CONTAINS(?author_name,?author_name2))  
}  
}  
}  
SERVICE<http://dblp.l3s.de/d2r/sparql>  
{  
  SELECT ?author3 ?author_name3 count(?publication) as ?num_publication  
  GROUP_CONCAT(?publication; SEPARATOR=", ") AS ?publications  
  WHERE  
  {  
    ?publication rdf:type swrc:Article;  
      dc:creator ?author3.  
    ?author3 rdfs:label ?author_name3.  
  }  
  GROUP BY ?author3 ?author_name3  
  ORDER BY DESC(?num_publication)  
  LIMIT 1000  
}  
}  
FILTER(CONTAINS(?author_name,?author_name3))  
}
```

# Querying with SPARQL

- Both Datasets : Query 5(Federated Query)
  - Finding Author Name, Publications, Num\_Publications (DBLP,LAK)
  - Results arranged acc. to Count (No. of LAK Publications)



```

PREFIX d2r: <http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/config.rdf#>
PREFIX swrc: <http://swrc.ontoware.org/ontology#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX map: <file:///home/diederich/d2r-server-0.3.2/dblp-mapping.n3#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX bibo:<http://purl.org/ontology/bibo/>

SELECT ?author_name ?organization ?num_article ?articles ?num_publication ?publications
WHERE
{
  SERVICE<http://meco.l3s.uni-hannover.de:8890/sparql>
  {
    SELECT ?author ?author_name ?organization ?num_article ?articles
    WHERE
    {
      GRAPH<http://lak-conference.org/papers>
      {
        SELECT ?author ?author_name count(?article) as ?num_article
        GROUP_CONCAT(?article; SEPARATOR=", ") AS ?articles
        WHERE
        {
          ?article a bibo:Article.
          ?article foaf:maker ?author.
          ?author rdfs:label ?author_name
        }GROUP BY ?author ?author_name
        ORDER BY DESC(?num_article)
        LIMIT 1000
      }
    }
  }
}
  
```

LAK Dataset

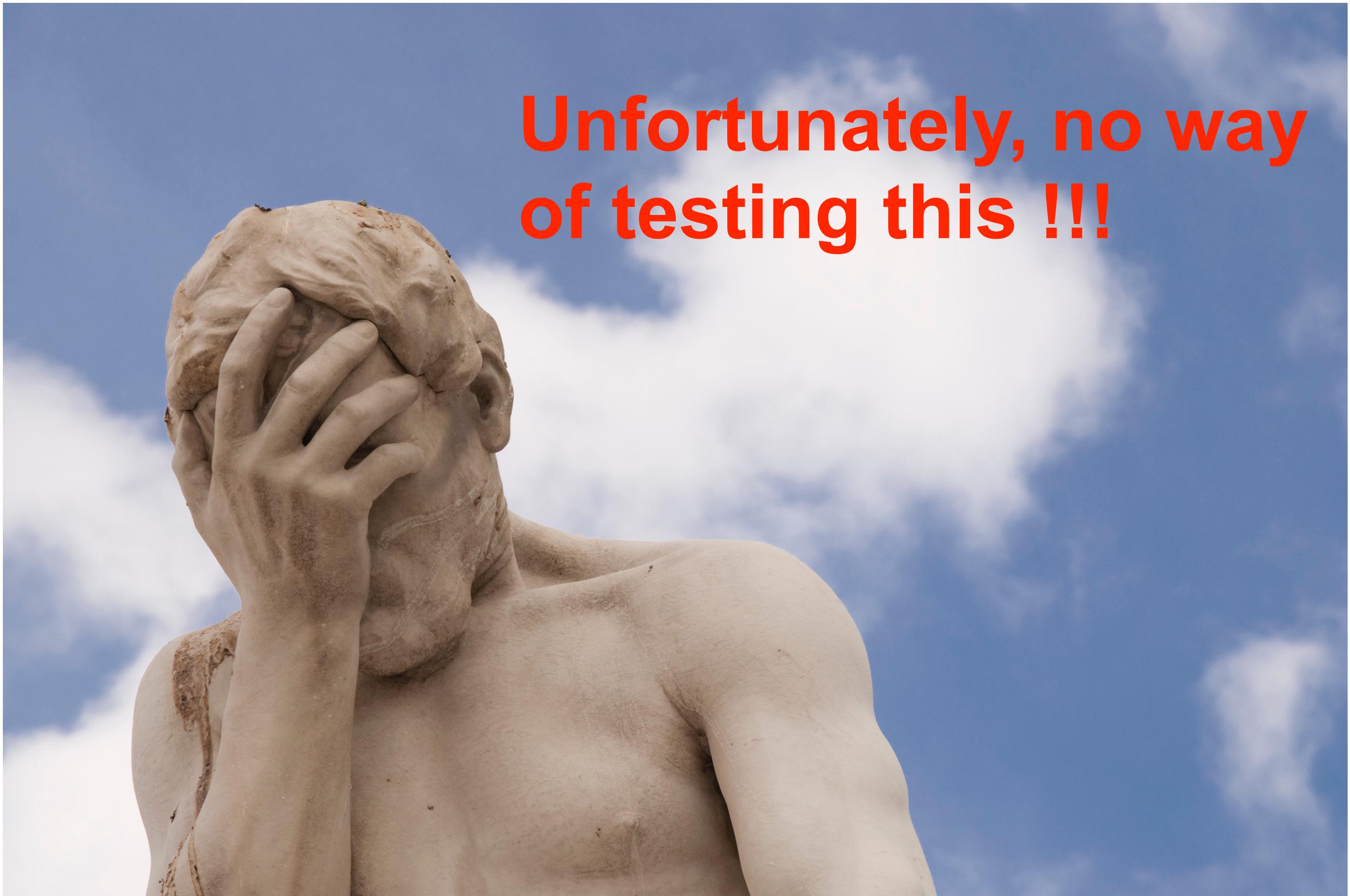
```

  {
    SELECT ?author2 ?author_name2 ?organization
    WHERE
    {
      ?author2 swrc:affiliation ?organization.
      ?author2 rdfs:label ?author_name2
    }
  }
  FILTER(CONTAINS(?author_name,?author_name2))
}
}
  
```

```

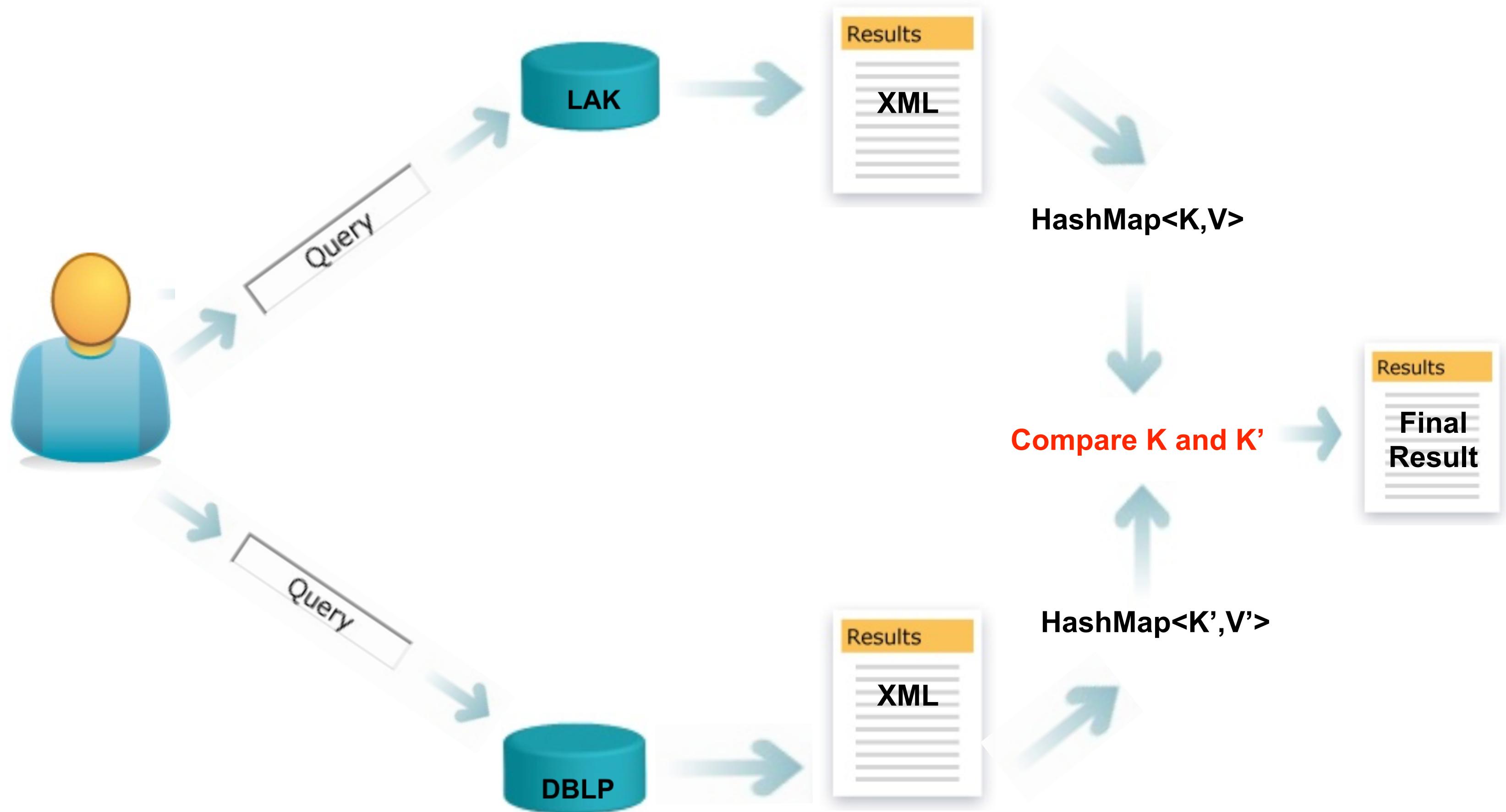
  SERVICE<http://dblp.l3s.de/d2r/sparql>
  {
    SELECT ?author3 ?author_name3 count(?publication) as ?num_publication
    GROUP_CONCAT(?publication; SEPARATOR=", ") AS ?publications
    WHERE
    {
      ?publication rdf:type swrc:Article;
      dc:creator ?author3.
      ?author3 rdfs:label ?author_name3.
    }GROUP BY ?author3 ?author_name3
    ORDER BY DESC(?num_publication)
    LIMIT 1000
  }
  FILTER(CONTAINS(?author_name,?author_name3))
}
  
```

DBLP Dataset



**Unfortunately, no way  
of testing this !!!**

## Second Approach



# Parsing XML File and Generating HashMap

- DBLP Dataset
  - XML File contains Author and Publication (obtained from Query A)
  - Generate DBLP\_HashMap<Author, Publication List>  
DBLP Dataset



**DBLP\_HashMap<Author, Publication List>**

<b>Key:Author</b>	<b>Value:Publication List</b>
Author 1	Publication A, Publication B
Author 2	Publication D
Author 3	Publication M, Pub X, Pub Z

# Parsing XML File and Generating HashMap

- LAK Dataset
  - XML File:Author, Org., Publication, Num\_Publication (Query 3)
  - Generate LAK\_HashMap<Author, LAK Details>



# Coreference Resolution

# Coreference Resolution



LAK\_HashMap<Author, Publication List>

Key: Author	Value: LAK Details
Author 3	Google, Publication List A, 10
Author 8	LUH, Publication List B, 7
Author 2	IBM, Publication List X, 3

# Coreference Resolution



LAK\_HashMap<Author, Publication List>

Key:Author	Value: LAK Details
Author 3	Google, Publication List A, 10
Author 8	LUH, Publication List B, 7
Author 2	IBM, Publication List X, 3



DBLP\_HashMap<Author, Publication List>

Key:Author	Value:Publication List
Author 1	Publication A, Publication B
Author 2	Publication D
Author 3	Publication M, Pub X, Pub Z

# Coreference Resolution



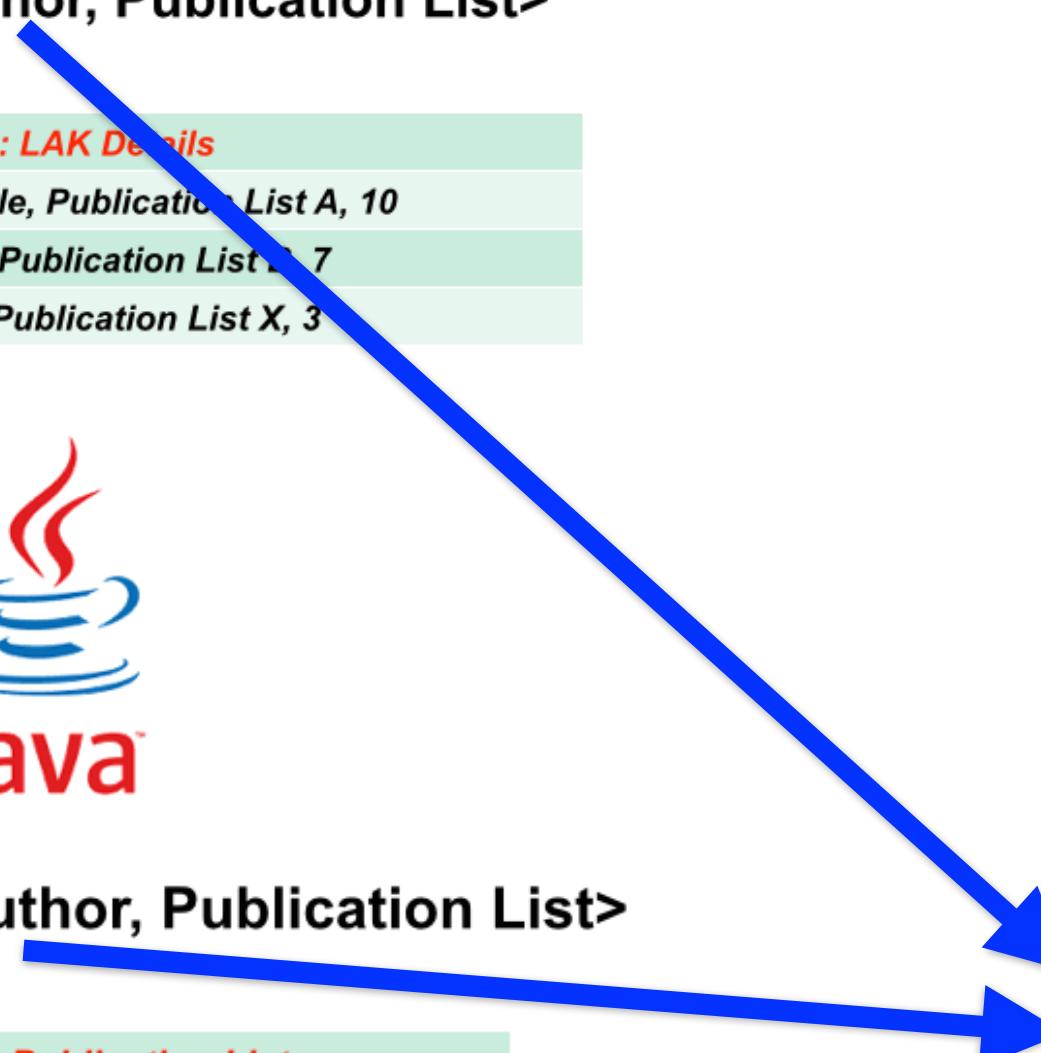
LAK\_HashMap<Author, Publication List>

Key:Author	Value: LAK Details
Author 3	Google, Publication List A, 10
Author 8	LUH, Publication List D, 7
Author 2	IBM, Publication List X, 3



DBLP\_HashMap<Author, Publication List>

Key:Author	Value:Publication List
Author 1	Publication A, Publication B
Author 2	Publication D
Author 3	Publication M, Pub X, Pub Z



# Coreference Resolution



**LAK\_HashMap<Author, Publication List>**

Key:Author	Value: LAK Details
Author 3	Google, Publication List A, 10
Author 8	LUH, Publication List D, 7
Author 2	IBM, Publication List X, 3



**DBLP\_HashMap<Author, Publication List>**

Key:Author	Value:Publication List
Author 1	Publication A, Publication B
Author 2	Publication D
Author 3	Publication M, Pub X, Pub Z

**Perfect Matches: String.equals() Method**  
**Imperfect Matches: Levenshtein Distance**

# Coreference Resolution



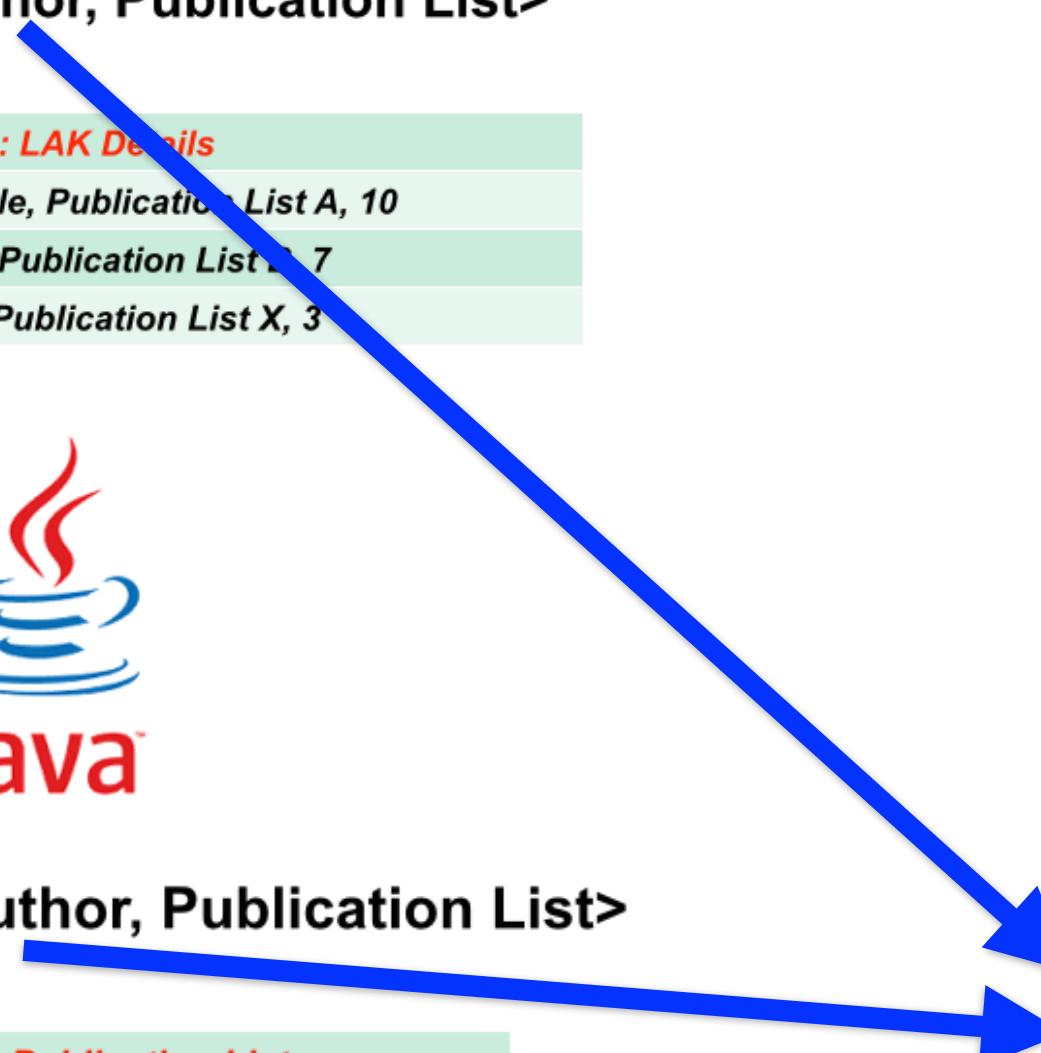
**LAK\_HashMap<Author, Publication List>**

Key:Author	Value: LAK Details
Author 3	Google, Publication List A, 10
Author 8	LUH, Publication List D, 7
Author 2	IBM, Publication List X, 3



**DBLP\_HashMap<Author, Publication List>**

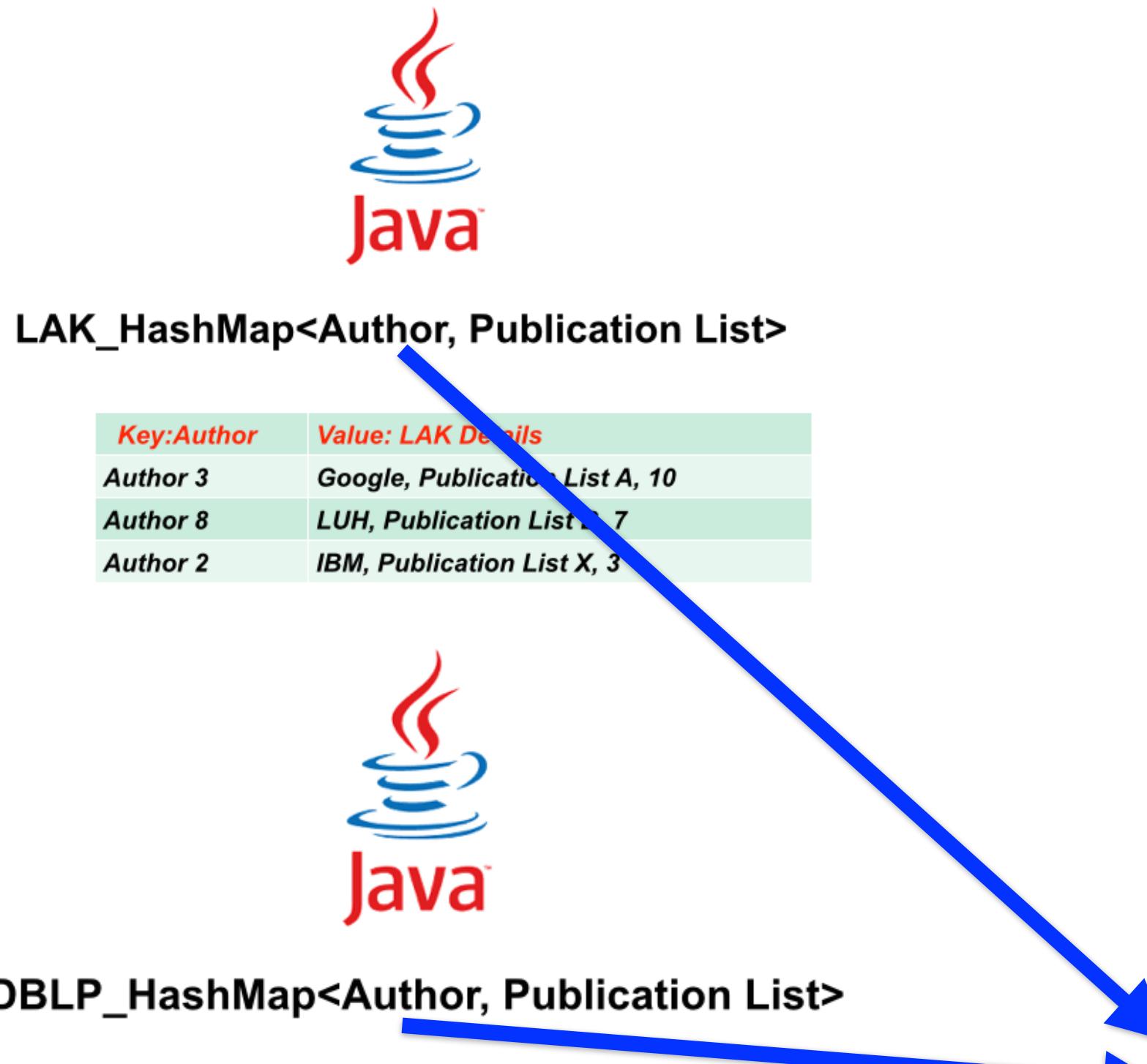
Key:Author	Value:Publication List
Author 1	Publication A, Publication B
Author 2	Publication D
Author 3	Publication M, Pub X, Pub Z



**Perfect Matches:** String.equals() Method  
**Imperfect Matches:** Levenshtein Distance

No. of Operations (Insert, Delete or Replace)  
 req. to convert one string to another

# Coreference Resolution



Author	LAK Details	DBLP Publications	No. of DBLP Publications
Author 3	Google, Publication List A, 10	Publication M, Pub X, Pub Z	3
Author 2	IBM, Publication List X, 3	Publication D	1

**Final Result**

**Perfect Matches:** String.equals() Method  
**Imperfect Matches:** Levenshtein Distance

**No. of Operations (Insert, Delete or Replace)**  
 req. to convert one string to another

# Coreference Resolution

# Coreference Resolution

File Path ▾ : ~/Desktop/project\_final/dblptry.xml

```

29      <literal>Istvan Beszteri</literal>
30    </binding>
31    <binding name="publication">
32      <uri>http://dblp.l3s.de/abc</uri>
33    </binding>
34  </result>
35  <result>
36    <binding name="author_name2">
37      <literal>TomCruise</literal>
38    </binding>
39    <binding name="publication">
40      <uri>Mission Impossible</uri>
41    </binding>
42  </result>
43  <result>
44    <binding name="author_name2">
45      <literal>TomCruise</literal>
46    </binding>
47    <binding name="publication">
48      <uri>Edge of Tomorrow</uri>
49    </binding>
50  </result>
51  <result>
52    <binding name="author_name2">
53      <literal>Robert De Niro</literal>
54    </binding>
55    <binding name="publication">
56      <uri>The Godfather</uri>
57    </binding>
58  </result>
59  <result>
60    <binding name="author_name2">
61      <literal>Robert De Niro</literal>
62    </binding>
63    <binding name="publication">
64      <uri>Silver Linings Playbook</uri>
65    </binding>
66  </result>
67 </results>

```

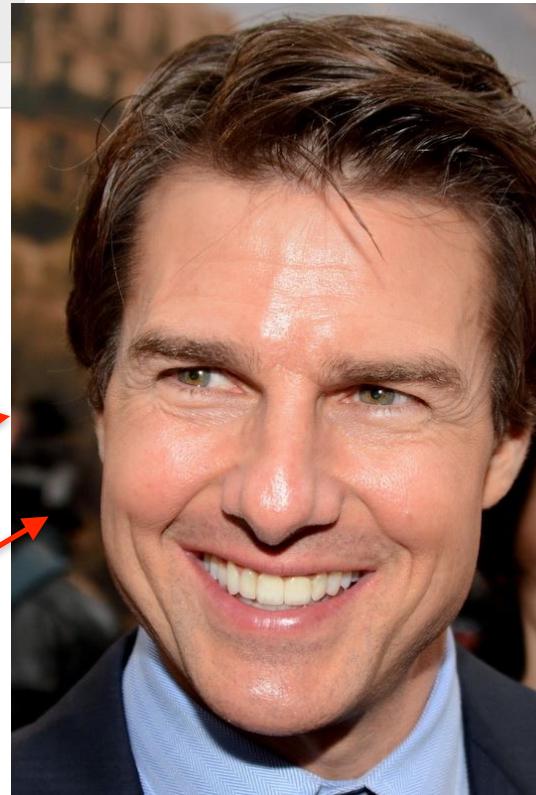
# Coreference Resolution

File Path ▾ : ~/Desktop/project\_final/dblptry.xml

```

29      <literal>Istvan Beszteri</literal>
30    </binding>
31    <binding name="publication">
32      <uri>http://dblp.l3s.de/abc</uri>
33    </binding>
34  </result>
35  <result>
36    <binding name="author_name2">
37      <literal>TomCruise</literal>
38    </binding>
39    <binding name="publication">
40      <uri>Mission Impossible</uri>
41    </binding>
42  </result>
43  <result>
44    <binding name="author_name2">
45      <literal>TomCruise</literal>
46    </binding>
47    <binding name="publication">
48      <uri>Edge of Tomorrow</uri>
49    </binding>
50  </result>
51  <result>
52    <binding name="author_name2">
53      <literal>Robert De Niro</literal>
54    </binding>
55    <binding name="publication">
56      <uri>The Godfather</uri>
57    </binding>
58  </result>
59  <result>
60    <binding name="author_name2">
61      <literal>Robert De Niro</literal>
62    </binding>
63    <binding name="publication">
64      <uri>Silver Linings Playbook</uri>
65    </binding>
66  </result>
67 </results>

```



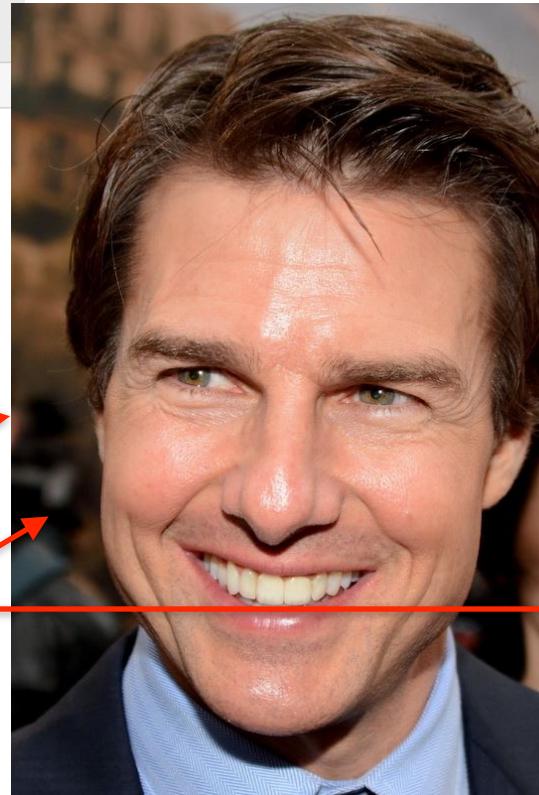
# Coreference Resolution

File Path ▾ : ~/Desktop/project\_final/dblptry.xml

```

29      <literal>Istvan Beszteri</literal>
30    </binding>
31    <binding name="publication">
32      <uri>http://dblp.l3s.de/abc</uri>
33    </binding>
34  </result>
35  <result>
36    <binding name="author_name2">
37      <literal>TomCruise</literal>
38    </binding>
39    <binding name="publication">
40      <uri>Mission Impossible</uri>
41    </binding>
42  </result>
43  <result>
44    <binding name="author_name2">
45      <literal>TomCruise</literal>
46    </binding>
47    <binding name="publication">
48      <uri>Edge of Tomorrow</uri>
49    </binding>
50  </result>
51  <result>
52    <binding name="author_name2">
53      <literal>Robert De Niro</literal>
54    </binding>
55    <binding name="publication">
56      <uri>The Godfather</uri>
57    </binding>
58  </result>
59  <result>
60    <binding name="author_name2">
61      <literal>Robert De Niro</literal>
62    </binding>
63    <binding name="publication">
64      <uri>Silver Linings Playbook</uri>
65    </binding>
66  </result>
67 </results>

```



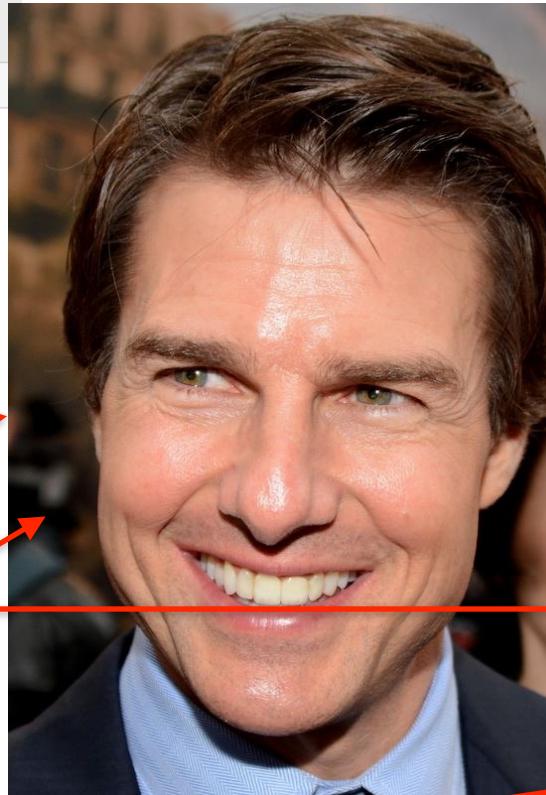
# Coreference Resolution

File Path ▾ : ~/Desktop/project\_final/dblptry.xml

```

29      <literal>Istvan Beszteri</literal>
30    </binding>
31    <binding name="publication">
32      <uri>http://dblp.l3s.de/abc</uri>
33    </binding>
34  </result>
35  <result>
36    <binding name="author_name2">
37      <literal>TomCruise</literal>
38    </binding>
39    <binding name="publication">
40      <uri>Mission Impossible</uri>
41    </binding>
42  </result>
43  <result>
44    <binding name="author_name2">
45      <literal>TomCruise</literal>
46    </binding>
47    <binding name="publication">
48      <uri>Edge of Tomorrow</uri>
49    </binding>
50  </result>
51  <result>
52    <binding name="author_name2">
53      <literal>Robert De Niro</literal>
54    </binding>
55    <binding name="publication">
56      <uri>The Godfather</uri>
57    </binding>
58  </result>
59  <result>
60    <binding name="author_name2">
61      <literal>Robert De Niro</literal>
62    </binding>
63    <binding name="publication">
64      <uri>Silver Linings Playbook</uri>
65    </binding>
66  </result>
67 </results>

```



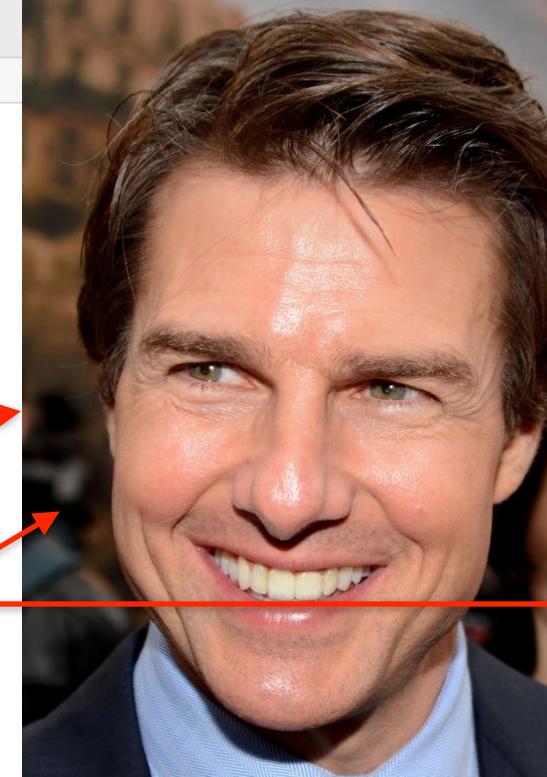
# Coreference Resolution

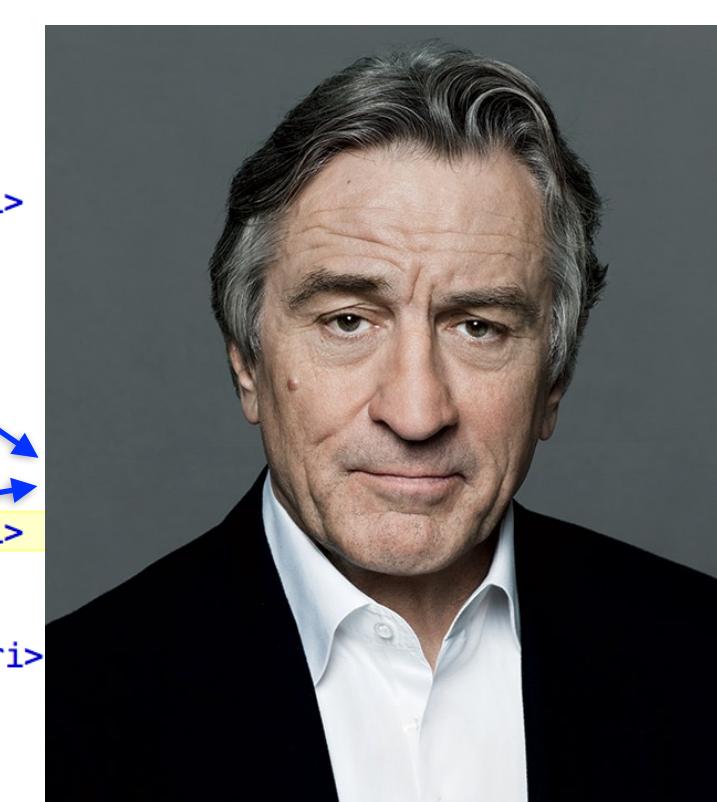
File Path ▾ : ~/Desktop/project\_final/dblptry.xml

```

29      <literal>Istvan Beszteri</literal>
30    </binding>
31    <binding name="publication">
32      <uri>http://dblp.l3s.de/abc</uri>
33    </binding>
34  </result>
35  <result>
36    <binding name="author_name2">
37      <literal>TomCruise</literal>
38    </binding>
39    <binding name="publication">
40      <uri>Mission Impossible</uri>
41    </binding>
42  </result>
43  <result>
44    <binding name="author_name2">
45      <literal>TomCruise</literal>
46    </binding>
47    <binding name="publication">
48      <uri>Edge of Tomorrow</uri>
49    </binding>
50  </result>
51  <result>
52    <binding name="author_name2">
53      <literal>Robert De Niro</literal>
54    </binding>
55    <binding name="publication">
56      <uri>The Godfather</uri>
57    </binding>
58  </result>
59  <result>
60    <binding name="author_name2">
61      <literal>Robert De Niro</literal>
62    </binding>
63    <binding name="publication">
64      <uri>Silver Linings Playbook</uri>
65    </binding>
66  </result>
67 </results>

```

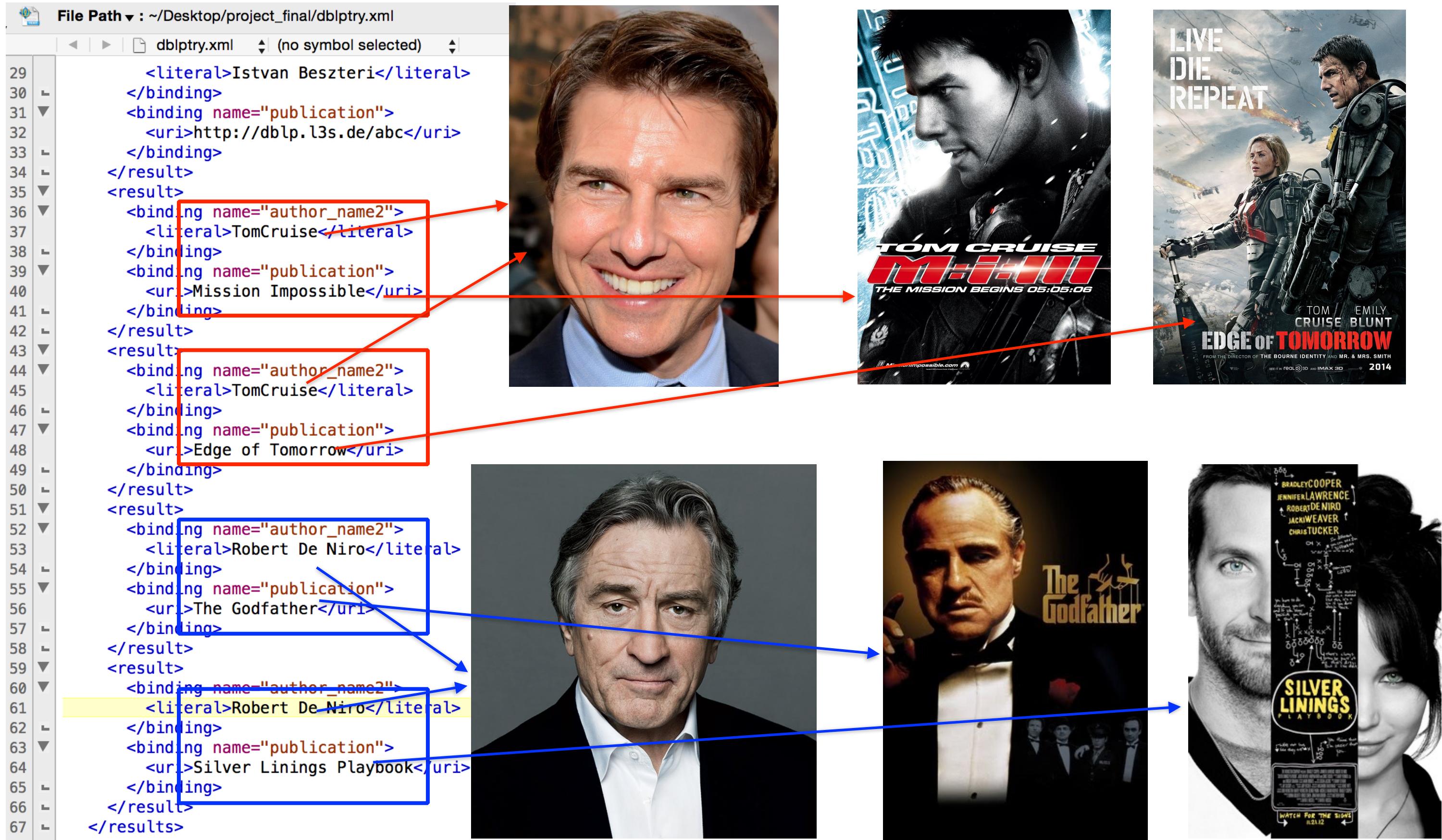


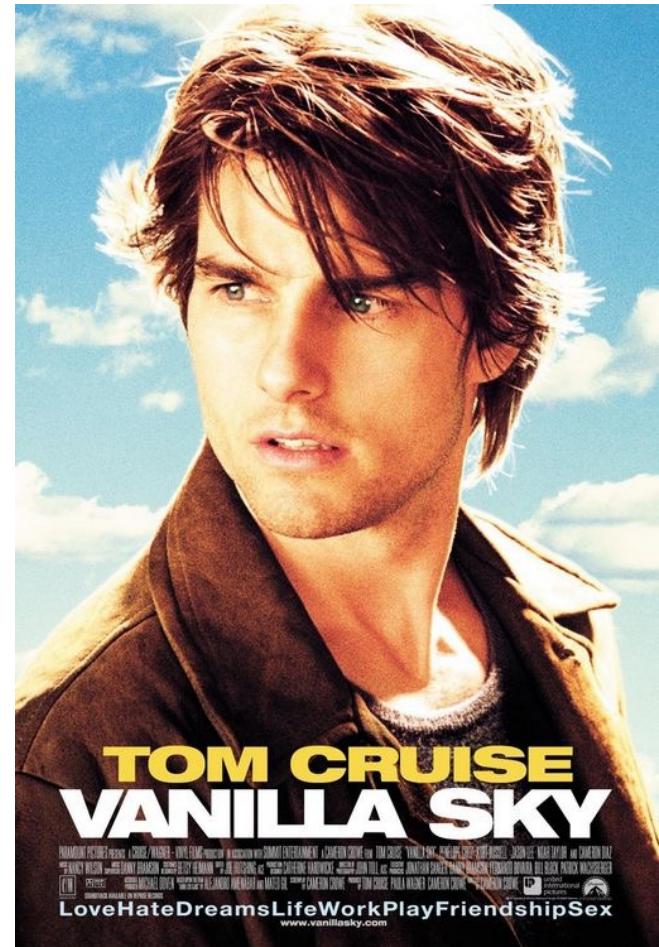



# Coreference Resolution



# Coreference Resolution





# Coreference Resolution

```

laktry.xml
ath ▼ : ~/Desktop/project_final/laktry.xml
▶ | □ laktry.xml (no symbol selected) ↻

ml version="1.0"?>
results>
result>
<binding name="author_name"><literal>Tiffany Barnes</literal></binding>
<binding name="organization_name"><literal>University of North Carolina at Charlotte</literal></binding>
<binding name="num_article"><literal datatype="http://www.w3.org/2001/XMLSchema#integer">17</literal></binding>
<binding name="articles"><literal>http://data.linkededucation.org/resource/lak/conference/edm2010/paper/175, http://da
/result>
result>
<binding name="author_name"><literal>Ryan S.j.d. Baker</literal></binding>
<binding name="organization_name"><literal>Worcester Polytechnic Institute</literal></binding>
<binding name="num_article"><literal datatype="http://www.w3.org/2001/XMLSchema#integer">40</literal></binding>
<binding name="articles"><literal>http://data.linkededucation.org/resource/lak/journal/jla201401/paper/479, http://data
</result>
<result>
<binding name="author_name"><literal>Neil T. Heffernan</literal></binding>
<binding name="organization_name"><literal>Worcester Polytechnic Institute</literal></binding>
<binding name="num_article"><literal datatype="http://www.w3.org/2001/XMLSchema#integer">25</literal></binding>
<binding name="articles"><literal>http://data.linkededucation.org/resource/lak/conference/edm2013/paper/241, http://da
</result>
<result>
<binding name="author_name"><literal>TomCruise</literal></binding>
<binding name="organization_name"><literal>Hollywood</literal></binding>
<binding name="num_article"><literal datatype="http://www.w3.org/2001/XMLSchema#integer">3</literal></binding>
<binding name="articles"><literal>Vanilla Sky, Jerry McGuire, Collateral</literal></binding>
</result>
<result>
<binding name="author_name"><literal>Robert D. NIR0</literal></binding>
<binding name="organization_name"><literal>OldHollywood</literal></binding>
<binding name="num_article"><literal datatype="http://www.w3.org/2001/XMLSchema#integer">2</literal></binding>
<binding name="articles"><literal>Dirty Grandpa, Goodfellas</literal></binding>
</result>
results>

```

# Coreference Resolution

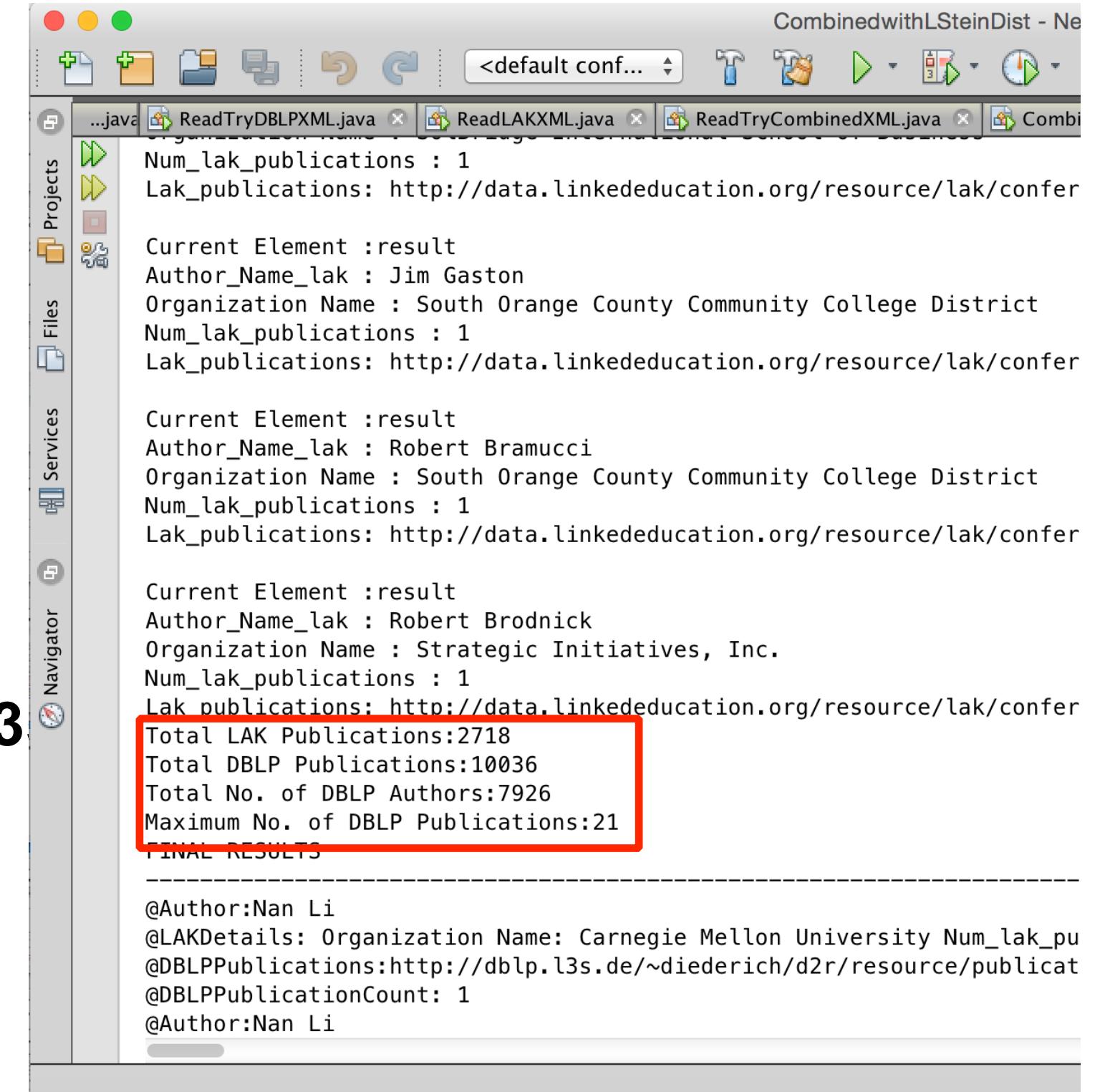
# Results

- **LAK Dataset**

- **No. of Results(Authors) = 1000**
- **Max. Publications = 40**
- **Min. Publications = 1**
- **Total Publications = 2718**
- **Avg. Publications = 2.72**

- **DBLP Dataset**

- **No. of Results(Publications) = 1003**
- **Total Authors = 7926**
- **Max. Publications = 21**
- **Min. Publications = 1**
- **Avg. Publications = 1.27**



```

CombinedwithLSteinDist - Ne
...java ReadTryDBLPXML.java ReadLAKXML.java ReadTryCombinedXML.java Combi
Projects Files Services Navigator
Num_lak_publications : 1
Lak_publications: http://data.linkededucation.org/resource/lak/confer
Current Element :result
Author_Name_lak : Jim Gaston
Organization Name : South Orange County Community College District
Num_lak_publications : 1
Lak_publications: http://data.linkededucation.org/resource/lak/confer

Current Element :result
Author_Name_lak : Robert Bramucci
Organization Name : South Orange County Community College District
Num_lak_publications : 1
Lak_publications: http://data.linkededucation.org/resource/lak/confer

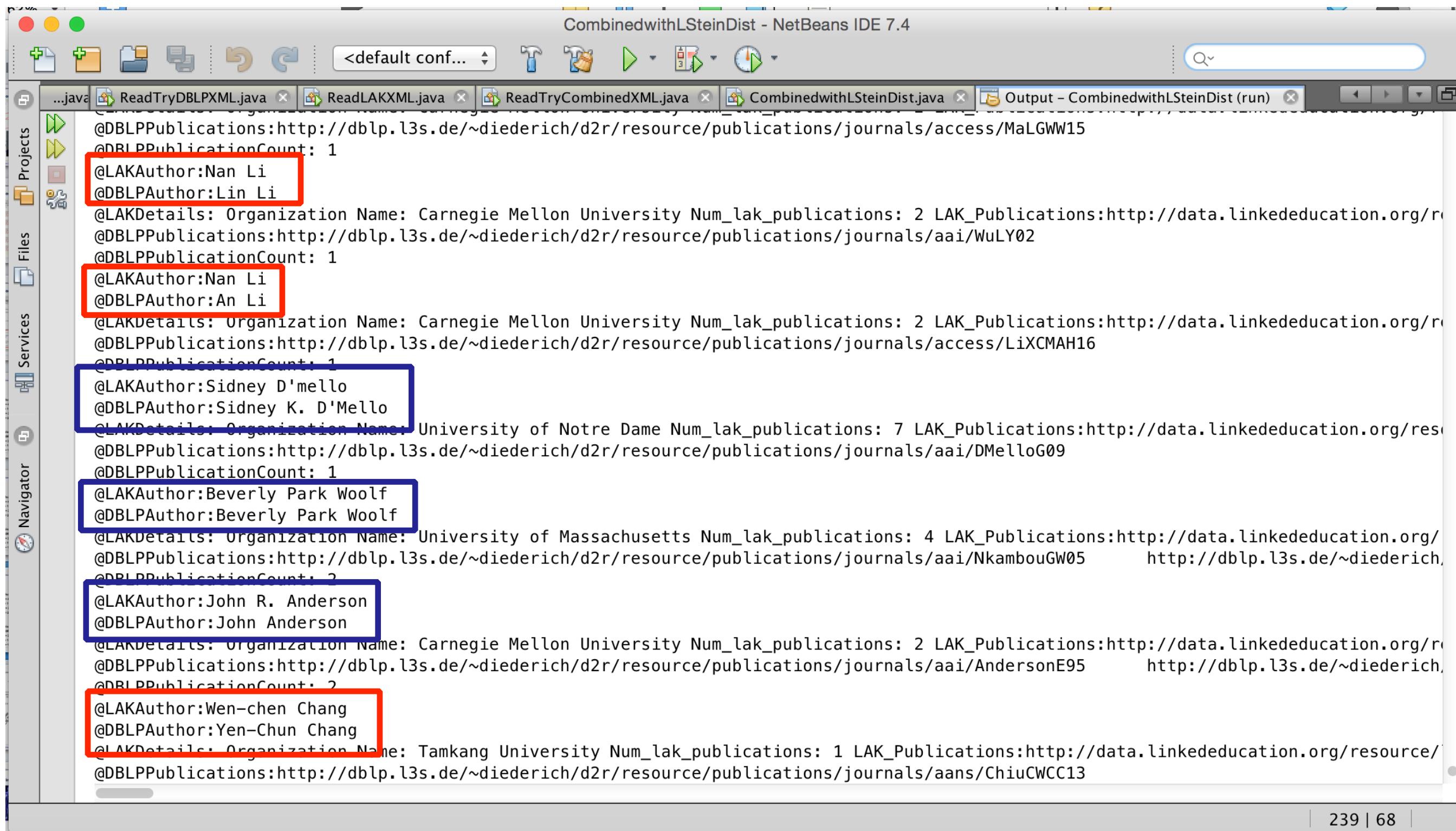
Current Element :result
Author_Name_lak : Robert Brodnick
Organization Name : Strategic Initiatives, Inc.
Num_lak_publications : 1
Lak_publications: http://data.linkededucation.org/resource/lak/confer
Total LAK Publications:2718
Total DBLP Publications:10036
Total No. of DBLP Authors:7926
Maximum No. of DBLP Publications:21
FINAL RESULTS

@Author:Nan Li
@LAKDetails: Organization Name: Carnegie Mellon University Num_lak_publ...
@DBLPPublications:http://dblp.l3s.de/~diederich/d2r/resource/publicat...
@DBLPPublicationCount: 1
@Author:Nan Li

```

# Evaluation

- Levenshtein Distance (<= 2)



The screenshot shows the NetBeans IDE interface with the title bar "CombinedwithLSteinDist - NetBeans IDE 7.4". The code editor displays Java code for reading XML files from DBLP and LAK datasets and comparing them using Levenshtein distance. The code includes annotations for DBLP publications, LAK author counts, and LAK details. Several names are highlighted with red and blue boxes to demonstrate the comparison logic.

```

CombinedwithLSteinDist - NetBeans IDE 7.4
<default config> Run Stop Output - CombinedwithLSteinDist (run)
...java ReadTryDBLPXML.java ReadLAKXML.java ReadTryCombinedXML.java CombinedwithLSteinDist.java Output - CombinedwithLSteinDist (run)

@DBLPublications:http://dblp.l3s.de/~diederich/d2r/resource/publications/journals/access/MaLGWW15
@DBLPublicationCount: 1
@LAKAuthor:Nan Li
@DBLPAuthor:Lin Li

@LAKDetails: Organization Name: Carnegie Mellon University Num_lak_publications: 2 LAK_Publications:http://data.linkededucation.org/resource/...
@DBLPublications:http://dblp.l3s.de/~diederich/d2r/resource/publications/journals/aaai/WuLY02
@DBLPublicationCount: 1
@LAKAuthor:Nan Li
@DBLPAuthor:An Li

@LAKDetails: Organization Name: Carnegie Mellon University Num_lak_publications: 2 LAK_Publications:http://data.linkededucation.org/resource/...
@DBLPublications:http://dblp.l3s.de/~diederich/d2r/resource/publications/journals/access/LiXCMAH16
@DBLPPublicationCount: 1
@LAKAuthor:Sidney D'mello
@DBLPAuthor:Sidney K. D'Mello
@LAKDetails: Organization Name: University of Notre Dame Num_lak_publications: 7 LAK_Publications:http://data.linkededucation.org/resource/...
@DBLPublications:http://dblp.l3s.de/~diederich/d2r/resource/publications/journals/aaai/DMelloG09
@DBLPPublicationCount: 1
@LAKAuthor:Beverly Park Woolf
@DBLPAuthor:Beverly Park Woolf

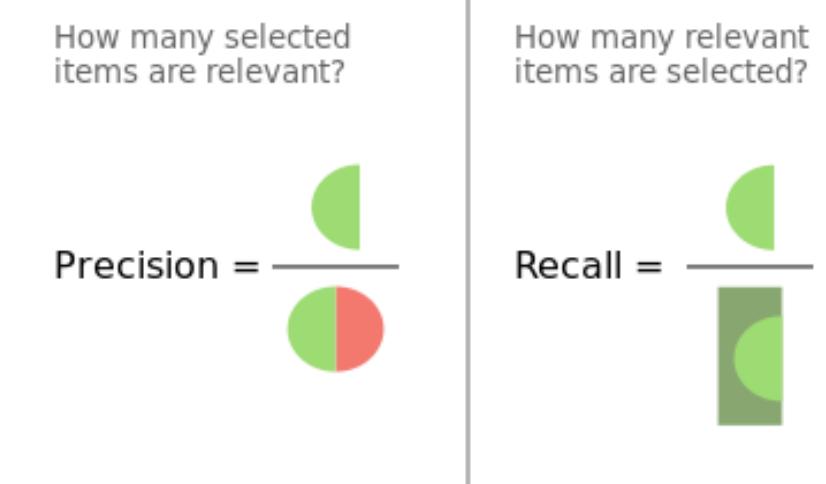
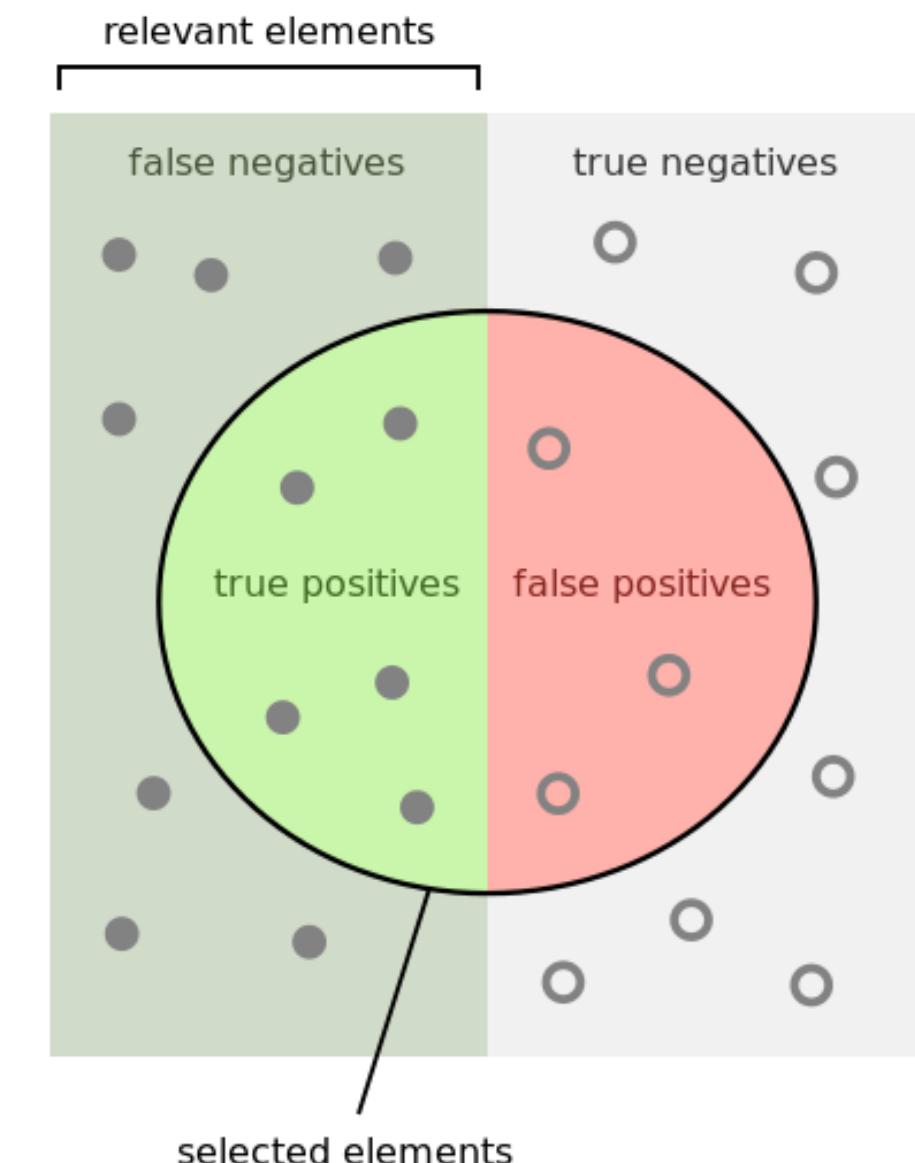
@LAKDetails: organization Name: University of Massachusetts Num_lak_publications: 4 LAK_Publications:http://data.linkededucation.org/resource/...
@DBLPublications:http://dblp.l3s.de/~diederich/d2r/resource/publications/journals/aaai/NkambouGW05 http://dblp.l3s.de/~diederich/d2r/resource/...
@DBLPPublicationCount: 2
@LAKAuthor:John R. Anderson
@DBLPAuthor:John Anderson

@LAKDetails: organization Name: Carnegie Mellon University Num_lak_publications: 2 LAK_Publications:http://data.linkededucation.org/resource/...
@DBLPublications:http://dblp.l3s.de/~diederich/d2r/resource/publications/journals/aaai/AndersonE95 http://dblp.l3s.de/~diederich/d2r/resource/...
@DBLPPublicationCount: 2
@LAKAuthor:Wen-chen Chang
@DBLPAuthor:Yen-Chun Chang
@LAKDetails: Organization Name: Tamkang University Num_lak_publications: 1 LAK_Publications:http://data.linkededucation.org/resource/...
@DBLPublications:http://dblp.l3s.de/~diederich/d2r/resource/publications/journals/aans/ChiuCWCC13

```

# Evaluation

- **Levensthein Distance (<2)**
  - Authors Matched = 29
  - Correct Matches = 19
  - (Perfect Matches = 16)
  - Precision =  $19/29 = 0.66$
  - Person(owl:sameAs) = 383
  - Recall =  $20/383 = 0.052$
- **Levensthein Distance (<3)**
  - Authors Matched = 100
  - Correct Matches = 22
  - Precision =  $22/100 = 0.22$
  - Person(owl:sameAs) = 383
  - Recall =  $22/383 = 0.055$



## Conclusion

- Support for AS needed in DBLP endpoint
- For Levenshtein Distance(<3): Precision greatly reduced
- Chinese Authors: Bad Results even keeping distance 1
- People with Middle Name: Good Results keeping distance upto 2

Lei Ye	Lei Yu
Lei Ye	Lei Yi
Lei Ye	Lei Hu
Lei Ye	Lei Li
Lei Ye	En Ye

Pamela Jordan	Pamela W. Jordan
Sidney D'Mello	Sidney K. D'Mello
Daniel L. Schwartz	Daniel Schwartz

- Limit on the Number of LAK Results (only 1000 fulfil all LAK Criteria)

## References

- [1]<http://www.mkyong.com/java/how-to-read-xml-file-in-java-dom-parser/>
- [2]<http://stackoverflow.com/questions/428073/what-is-the-best-simplest-way-to-read-in-an-xml-file-in-java-application>
- [3]<http://codereview.stackexchange.com/questions/24116/hash-table-implementation-in-java>
- [4]<http://stackoverflow.com/questions/1066589/iterate-through-a-hashmap>
- [5]<http://stackoverflow.com/questions/23143326/how-can-i-add-a-value-to-an-array-by-referencing-the-hashmap-containing-it-in-ja?rq=1>
- [6]<http://stackoverflow.com/questions/599161/best-way-to-convert-an-arraylist-to-a-string>
- [7]<http://stackoverflow.com/questions/10462819/get-keys-from-hashmap-in-java>
- [8]<http://www.java2s.com/Code/Java/Data-Type/FindtheLevenshteindistancebetweentwoStrings.htm>

# Discussion

