

<http://eprints.ecs.soton.ac.uk/id/eprint/21897>

Co-reference and sameAs.org

DevCSI Linked Data Hackdays
ILRT, Bristol, UK, 14th January 2011

Hugh Glaser
Seme4 Limited and University of Southampton, UK

Seme4
18 Soho Square
London W1 3QL
T : + 44 20 7060 1590

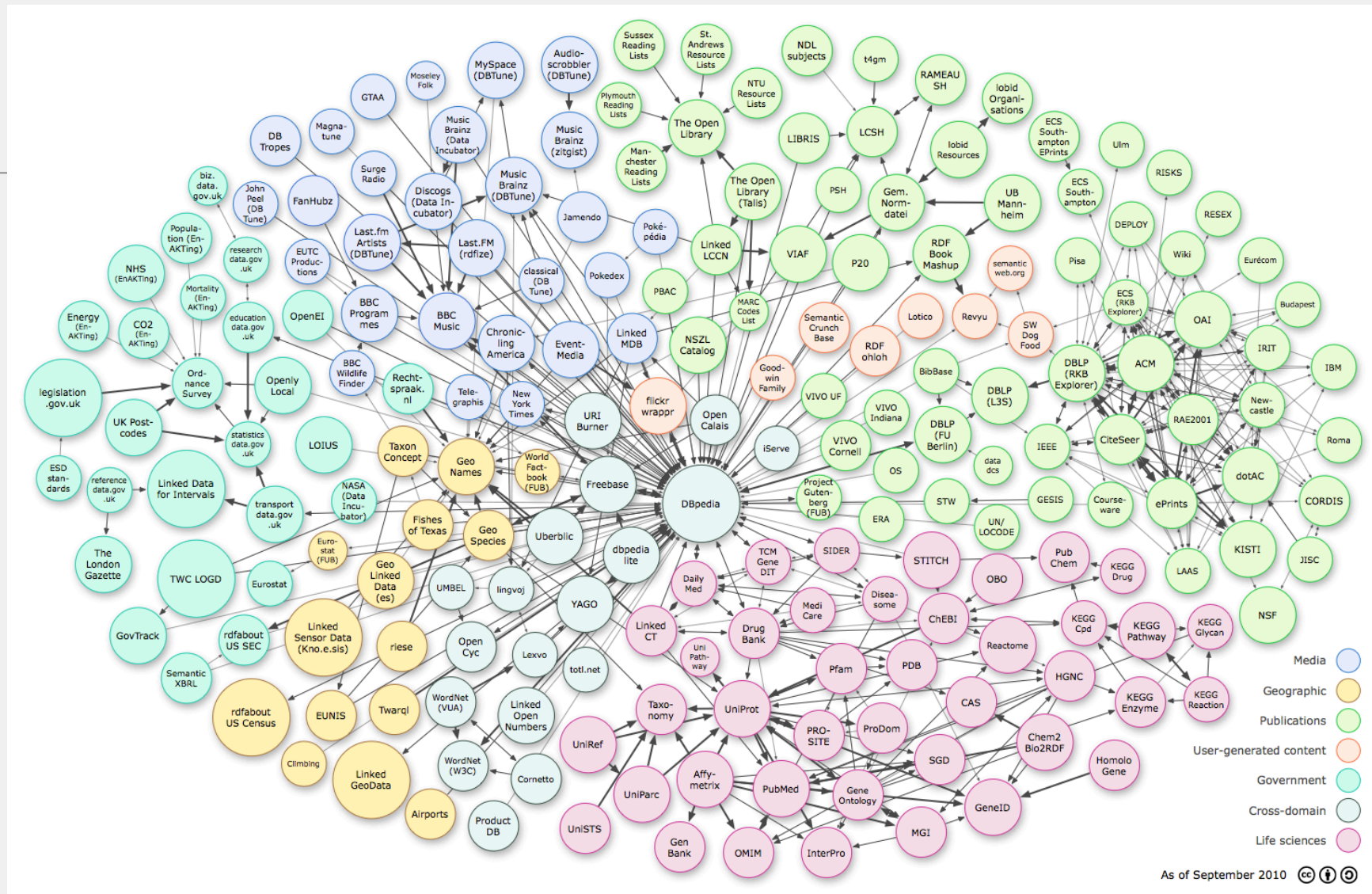
© 2010 Seme4

UNIVERSITY OF
Southampton
School of Electronics
and Computer Science



Seme4.com

The LOD “cloud”



Co-Reference

- **Lots of resources with multiple URIs**
 - This is the nature of the Beast
 - Legal, Sociological and Technical reasons
- **This is a Big Problem**
 - Everything is a URI (not title, name, number...)
 - Identifying multiple URIs for one resource
 - Rejecting incorrectly conflated resources
 - Publishing
 - Using
- **Solution**
 - Co-reference is just Knowledge, but quite special, and crucial, knowledge
 - Generating a new URI is not the solution to already having too many
 - The web is anarchic – don't try to impose impossible structure on it
 - Embrace the anarchy and multiple URIs in your systems ab initio

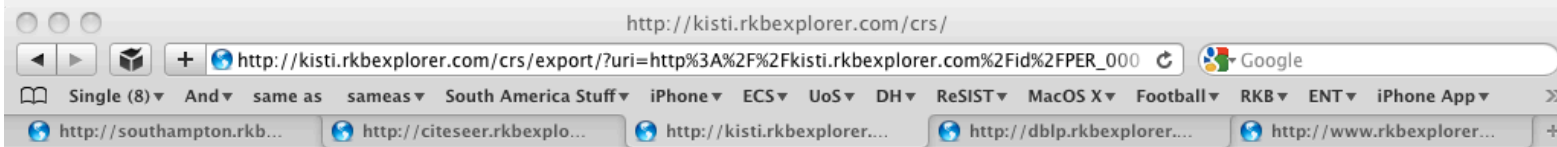
Co-Reference Service (CRS)

- **CRS Subsystem**
 - Find co-references
 - Store them
 - Publish them
 - Essentially:
 - $URI_i \rightarrow \{ URI_1, \dots, URI_i, \dots, URI_n \}$
 - Recommend a “Canon”
- **Published by the Data Publisher**
 - And possibly others
- **Middleware aggregates co-references from recognised CRSes**

CRS continued

- **CRS Policies are defined by context**
 - Often one per Triplestore
 - Can be many per Triplestore for different purposes
 - May not be associated with a particular Triplestore
- **Maintenance**
 - Provenance
 - Rollback
- **Can be used to infer owl:sameAs**
- **Eg OAI CRS has**
 - 7531045 different URIs
 - in
 - 2544955 bundles

Look up Seungwoo



kisti.rkbexplorer.com

[search](#) [query](#) [crs](#) [contact](#)

Canon

http://kisti.rkbexplorer.com/id/PER_0000000000000131417

Duplicate URIs

<http://acm.rkbexplorer.com/id/person-407157-c23d3b44f74d68cce281e3ff7795bf4c>

<http://citeseer.rkbexplorer.com/id/resource-CSP180445-4e6f59a7e2af5838836a83e27748b7ac>

<http://dblp.rkbexplorer.com/id/people-1ec5a600299222dd6374695ef5214f05-3aac82bb29cfbecfc2e83d8b23349445>

http://kisti.rkbexplorer.com/id/PER_0000000000000131417

Bundle instantiated at

2009-12-20 02:39:10

Alternative representations

[RDF/XML](#)

[sameAs list in N3](#)

[image/png](#)



Seme4.com

Co-Reference Closure

Complete Co-Reference Information

This service computes the equivalence class within the known URIs for a specified URI, by consulting all relevant CRS knowledge bases.

Go

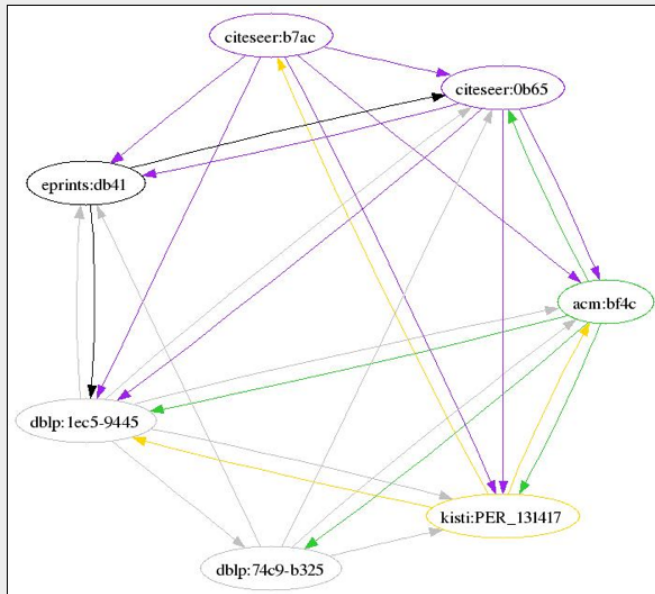
Display all related triples? ☐

Display connectivity image? ☒ None ☐ JPEG ☐ PNG ☐ GIF ☐ Postscript ☐ SVG

Equivalent URIs...

1. **(Canon)** <http://acm.rkbexplorer.com/id/person-407157-c23d3b44f74d68cce281e3ff7795bf4c>
2. <http://citeseer.rkbexplorer.com/id/resource-CSP180445-4e6f59a7e2af5838836a83e27748b7ac>
3. <http://citeseer.rkbexplorer.com/id/resource-CSP180445-e8a3df831f61dc5892835f5sec100b65>
4. <http://dblp.rkbexplorer.com/id/people-1ec5a600299222dd6374695ef5214f05-3aac82bb29cfbecf2e83d8b23349445>
5. <http://dblp.rkbexplorer.com/id/people-74c9b4bc601253864bbb6e2f83244264-b4da84ca0a3d5fe19b5c36c45fe3b325>
6. <http://eprints.rkbexplorer.com/id/kfupm/person-1ec5a600299222dd6374695ef5214f05-5c14713b5937e6d31a4714469e60db41>
7. http://kisti.rkbexplorer.com/id/PER_0000000000000131417

The following diagram shows the interconnectivity between the CRS knowledge bases which maintain the context-dependent representation of coreference for each of the RKBExplorer domains.



<http://acm.rkbexplorer.com/id/person-407157-c23d3b44f74d68cce281e3ff7795bf4c>

Showing information queried from all repositories ...

Subject	Property	Object/Value	Source
Seungwoo Lee	akt:full-name	Seungwoo Lee [Explore]	acm-periodicals.rdf >>
Seungwoo Lee	akt:full-name	Seungwoo Lee [Explore]	acm-proceedings.rdf >>
Seungwoo Lee	akt:full-name	Seungwoo Lee [Explore]	dblp-publications-2001.rdf >>
Seungwoo Lee	akt:has-affiliation	Electrical and Computer Engineering Division, Pohang University of Science & Technology (POSTECH), Pohang, South Korea. gblee@postech.ac.kr	acm-periodicals.rdf >>
Seungwoo Lee	akt:has-affiliation	POSTECH, Pohang, Korea	acm-proceedings.rdf >>
Seungwoo Lee	kisti:engNameOfPerson	Seungwoo Lee [Explore]	datatypeproperties.ttl >
Seungwoo Lee	rdf:type	akt:Affiliated-Person	acm-periodicals.rdf >>
Seungwoo Lee	rdf:type	akt:Affiliated-Person	acm-proceedings.rdf >>
Seungwoo Lee	rdf:type	Generic Agent	acm-periodicals.rdf >>
Seungwoo Lee	rdf:type	Generic Agent	acm-proceedings.rdf >>
Seungwoo Lee	rdf:type	Generic Agent	dblp-publications-2001.rdf >>
Seungwoo Lee	rdf:type	akt:Person	acm-periodicals.rdf >>
Seungwoo Lee	rdf:type	akt:Person	acm-proceedings.rdf >>
Seungwoo Lee	rdf:type	akt:Person	dblp-publications-2001.rdf >>
Seungwoo Lee	rdf:type	PER_char(20)^^^	datatypeproperties.ttl >
Seungwoo Lee	rdf:type	PER_char(20)^^^	objectproperties.ttl >>
Seungwoo Lee	rdf:type	PER_char(20)^^^	resources.ttl >>

Subject	Property	Object	Source
Automatic acquisition of named entity tagged corpus from world wide web	akt:has-author	Seungwoo Lee	acm-proceedings.rdf >>
A Corpus-Based Learning Method of Compound Noun Indexing Rules for Korean	akt:has-author	Seungwoo Lee	acm-periodicals.rdf >>
SiteQ: Engineering High Performance QA System Using Lexico-Semantic Pattern Matching and Shallow NLP.	akt:has-author	Seungwoo Lee	dblp-publications-2001.rdf >>
A Corpus-Based Learning Method of Compound Noun			dblp-publications-2001.rdf >>



LINKED DATA SOLUTIONS

Seme4.com

sameAs

http://sameas.org/ Google

Single (8) And same as sameas South America Stuff iPhone ECS UoS DH ReSIST MacOS X Football RKB ENT iPhone App

<sameAs>

interlinking the Web of Data

The Web of Data has many equivalent URIs.
This service helps you to find co-references
between different data sets.
Enter a known URI, or use Sindice to search first.


<sameAs> Enter a Linked Data URI...

Sindice Enter a literal (text) search...

Why not try searching for the string "[Southampton](#)", (which we will look up for you on [Sindice](#) first)
or finding other equivalent identifiers for <http://transport.data.gov.uk/id/local-authority/1755>?

Currently serving **39 221 073** URIs in **13 112 833** bundles!


about · contact · get the widget



 CC BY-NC-SA

<sameAs>

interlinking the Web of Data

The Web of Data has many equivalent URIs.
This service helps you to find co-references
between different data sets.
Enter a known URI, or use Sindice to search first.

<sameAs> 

Equivalent URIs for http://acm.rkbexplorer.com/id/person-344165-529a2367796b9822d610a72843a3980c –

<sameAs>

1. <http://data.semanticweb.org/person/atanas-kiryakov>
2. <http://acm.rkbexplorer.com/id/person-344165-529a2367796b9822d610a72843a3980c>
3. <http://acm.rkbexplorer.com/id/person-344165-be4c940af0b483a2d1675df9fa0bab6f>
4. <http://acm.rkbexplorer.com/id/person-344165-c23ba3f792faeca763400f197b89adb1>
5. <http://acm.rkbexplorer.com/id/person-344165-e0b9eea0f283efab1c5ae86ad61bf111>
6. <http://acm.rkbexplorer.com/id/person-344165-fa8e9aed948670c2accad97a4320122d>
7. <http://dblp.rkbexplorer.com/id/people-2ceb8136f69a235014d3d46c881a1...86ce3c9dc7>
8. <http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...f7185e51d3>
9. <http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...88ed85ab7e>
10. <http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...601d0b0236>
11. <http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...b93fb59e9f>
12. <http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...872498927e>
13. <http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...f41b54ff75>
14. <http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...3621e49944>
15. <http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...3a2580ac89>
16. <http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...b8848ff4e0>
17. <http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...6922936d8c>
18. <http://dblp.rkbexplorer.com/id/people-c9d93beb4ba98fa7318516e5fb5f9...c42f7e7ace>
19. <http://citeseer.rkbexplorer.com/id/resource-CSP162937-13aaf289f2945...80cea8e74c>
20. <http://citeseer.rkbexplorer.com/id/resource-CSP162937-2897e9670649a...00d1d4c62c>
21. <http://citeseer.rkbexplorer.com/id/resource-CSP162937-4bfe00e2ed7ef...904f25f2f2>
22. <http://citeseer.rkbexplorer.com/id/resource-CSP162937-64595ee407c94...7867a8505d>
23. <http://citeseer.rkbexplorer.com/id/resource-CSP162937-6b4779db34ec9...afd234cdc4>
24. <http://citeseer.rkbexplorer.com/id/resource-CSP162937-7896446de27c6...8f39db880b>
25. <http://citeseer.rkbexplorer.com/id/resource-CSP162937-8ac5adb3359f3...213751b546>

What is Southampton?

sameAs « http://transport.data.gov.uk/id/local...ority/1755

« http://sameas.org/html?uri=http%3A%2F%2Ftransport.data.gov.uk%2Fid%2Flocal-authority%2F1755 » Google

Single (8) ▾ And ▾ same as ▾ sameas ▾ South America Stuff ▾ iPhone ▾ ECS ▾ UoS ▾ DH ▾ ReSIST ▾ MacOS X ▾ Football ▾ RKB ▾ ENT ▾ iPhone App ▾ »

sameAs « http://transport.data.g...

<sameAs>

13. http://dbpedia.org/resource/Southampton_City_Centre
14. http://dbpedia.org/resource/Southampton,_England
15. http://dbpedia.org/resource/Southampton,_Hampshire
16. http://dbpedia.org/resource/Southampton_City_Council
17. [http://dbpedia.org/resource/Southampton_\(district\)](http://dbpedia.org/resource/Southampton_(district))
18. <http://www.okkam.org/ens/id13973975-c3dd-47e7-8e35-f09428d8723c>
19. <http://unlocode.rkbexplorer.com/id/GBSOU>
20. <http://unlocode.rkbexplorer.com/description/GBSOU>
21. <http://os.rkbexplorer.com/id/osr7000000000037256>
22. <http://d.opencalais.com/en/geo/city/ralg-geo1/8687634d-b732-c7f1-b3...3194e82d15>
23. <http://data.nytimes.com/N8998838277808400221>
24. http://data.nytimes.com/southampton_england_geo
25. <http://data.ordnancesurvey.co.uk/id/7000000000037256>
26. <http://linkedgeodata.org/triify/node/285832#id>
27. <http://mortality.psi.enakting.org/id/Southampton>
28. <http://mpii.de/yago/resource/Southampton>
29. <http://nuts.psi.enakting.org/id/UKJ32>
30. <http://openlylocal.com/id/councils/377>
31. <http://rdf.freebase.com/ns/en.southampton>
32. <http://rdf.freebase.com/ns/guid.9202a8c04000641f800000000075d87>
33. <http://rdf.freebase.com/ns/guid.9202a8c04000641f80000000004825363>
34. <http://revyu.com/tags/southampton>
35. <http://statistics.data.gov.uk/id/local-authority/00MS>
36. <http://statistics.data.gov.uk/id/local-authority/24UM>
37. <http://statistics.data.gov.uk/id/county?name=Southampton>
38. <http://statistics.data.gov.uk/id/local-authority-district/00MS>
39. <http://statistics.data.gov.uk/id/local-authority-district?name=Southampton>
40. <http://statistics.data.gov.uk/id/local-authority?name=Southampton City Council>
41. <http://statistics.data.gov.uk/id/local-authority-district?name=The ...outhampton>
42. <http://sw.cyc.com/concept/Mx4rvViRB5wpEbGdrcN5Y29ycA>
43. <http://sw.opencyc.org/concept/Mx4rvViRB5wpEbGdrcN5Y29ycA>
44. <http://sw.opencyc.org/concept/Mx4rvx6ncZwpEbGdrcN5Y29ycA>
45. <http://sw.opencyc.org/2008/06/10/concept/Mx4rvViRB5wpEbGdrcN5Y29ycA>
46. <http://sw.opencyc.org/2008/06/10/concept/Mx4rvx6ncZwpEbGdrcN5Y29ycA>
47. <http://sw.opencyc.org/2009/04/07/concept/Mx4rvViRB5wpEbGdrcN5Y29ycA>
48. <http://sw.opencyc.org/2009/04/07/concept/Mx4rvx6ncZwpEbGdrcN5Y29ycA>
49. <http://sw.opencyc.org/2008/06/10/concept/en/CityOfSouthamptonEngland>
50. <http://sw.opencyc.org/2009/04/07/concept/en/CityOfSouthamptonEngland>
51. http://sw.opencyc.org/2008/06/10/concept/en/Southampton_ParishBermuda
52. http://sw.opencyc.org/2009/04/07/concept/en/Southampton_ParishBermuda
53. <http://sws.geonames.org/2637487/>
54. <http://transport.data.gov.uk/id/local-authority/1755>
55. <http://transport.data.gov.uk/id/local-authority-district/1755>
56. <http://umbel.org/umbel/ne/wikipedia/Southampton>
57. <http://www.twine.com/item/1r2lcxqy-r1z>

rdf+xml · n3 · json · text · show fewer items

Seme4
LINKED DATA SOLUTIONS
Seme4.com

Concluding Remarks

- **RKBExplorer uses CRS services successfully**
 - Often 100s of URIs per person
- **British Museum uses a CRS service for Data Fusion on its web site**
- **Organisations can go ahead and build systems, then do the identity management**
 - Both internally and externally
 - Doing identity management first requires cross-organisation agreement, a delay of months to years
 - It then freezes the business processes to conform to the external requirement
- **sameAs.org turns out to be pretty useful**
 - > 10K hits/day
- **Build systems that accept the way the world is, not what you would like it to be**

You are never alone

- **Ian Millard**
- **ReSIST Project**
- **AKT Project**
- **EnAKTing Project**
- **10 years of collaborators**

<http://eprints.ecs.soton.ac.uk/id/eprint/21897>