

Discovering Concept Coverings in Aligning Ontologies of Linked Data

Rahul Parundekar, Craig A. Knoblock and Jose-Luis Ambite

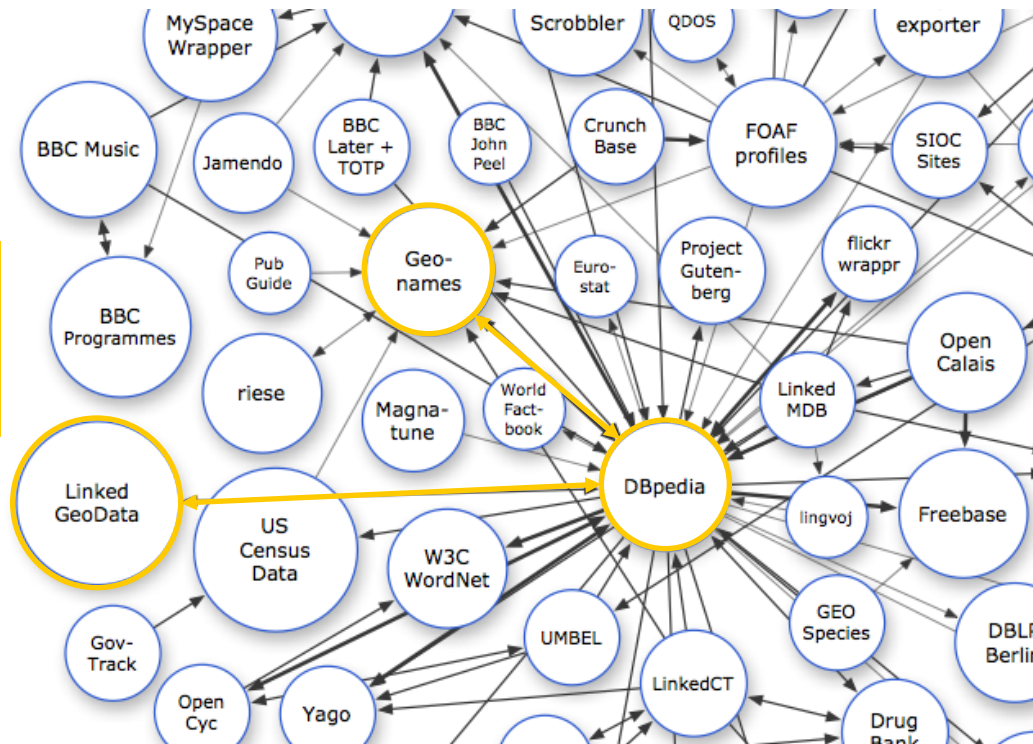
{parundek,knoblock}@usc.edu, ambite@isi.edu

University of Southern California

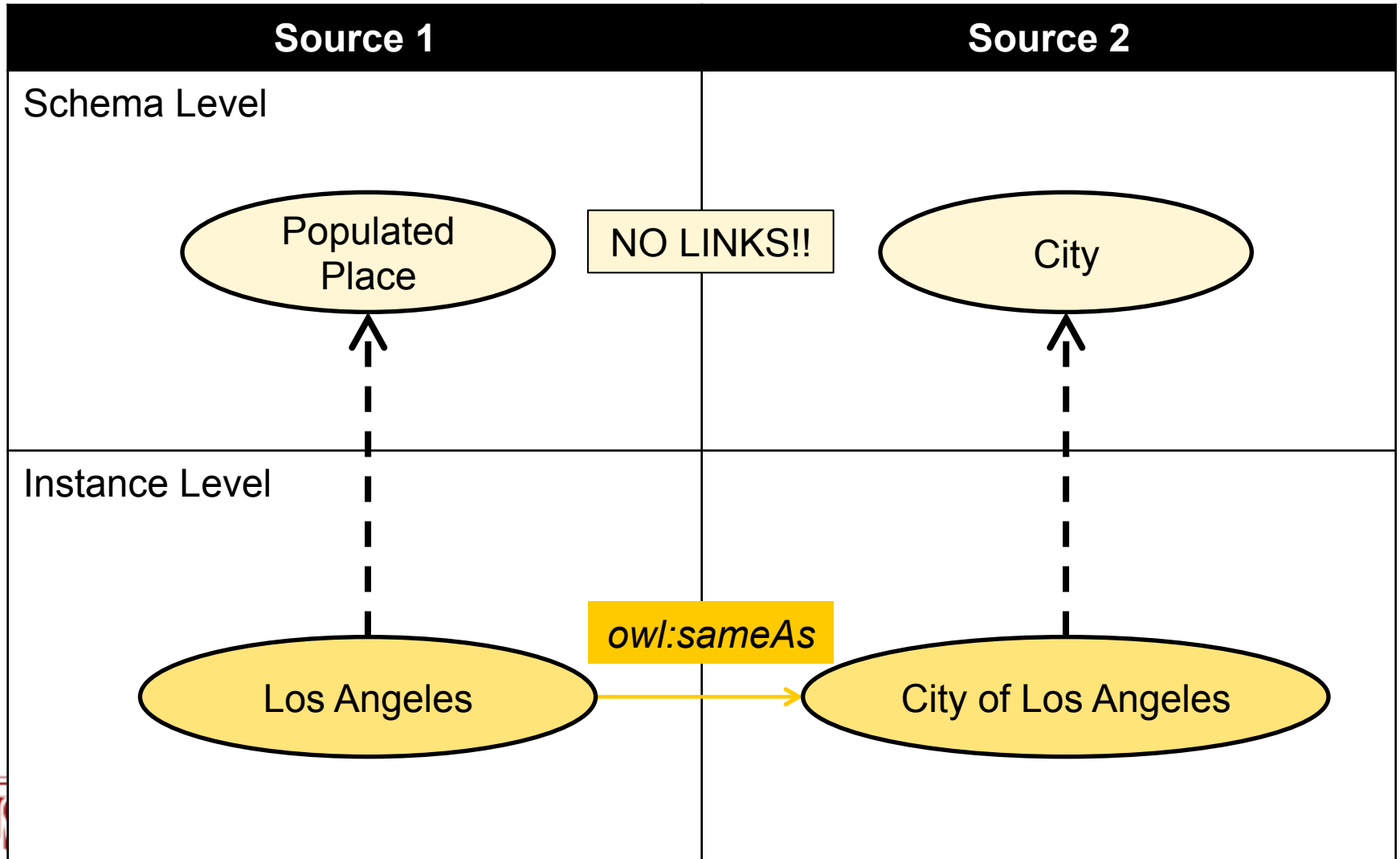
Web of Linked Data

- Different sources with different schemas
- Equivalent instances in the different domains connected with *owl:sameAs*

Example:
Geospatial
Domain





Interlinked instances with disjoint schemas

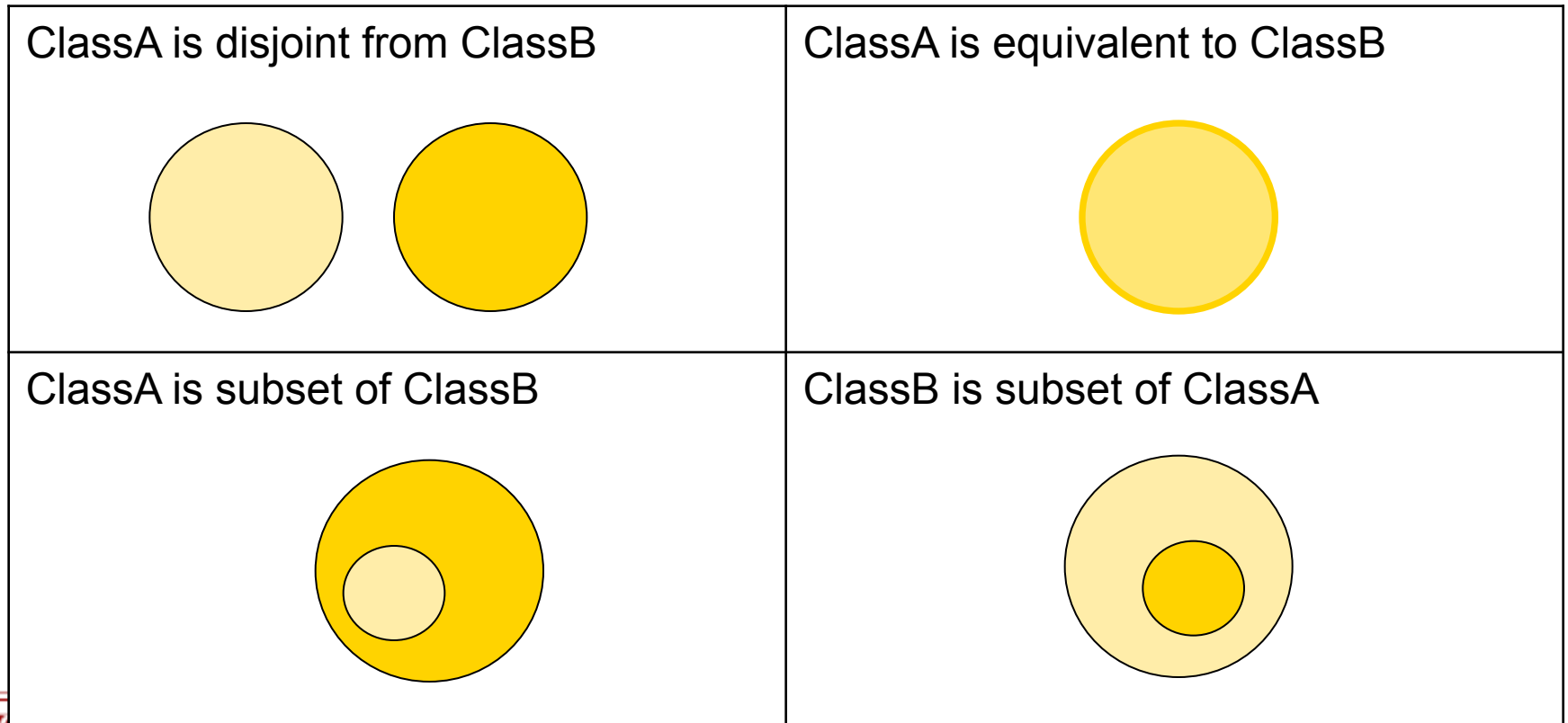


Problem: Disconnected Ontologies

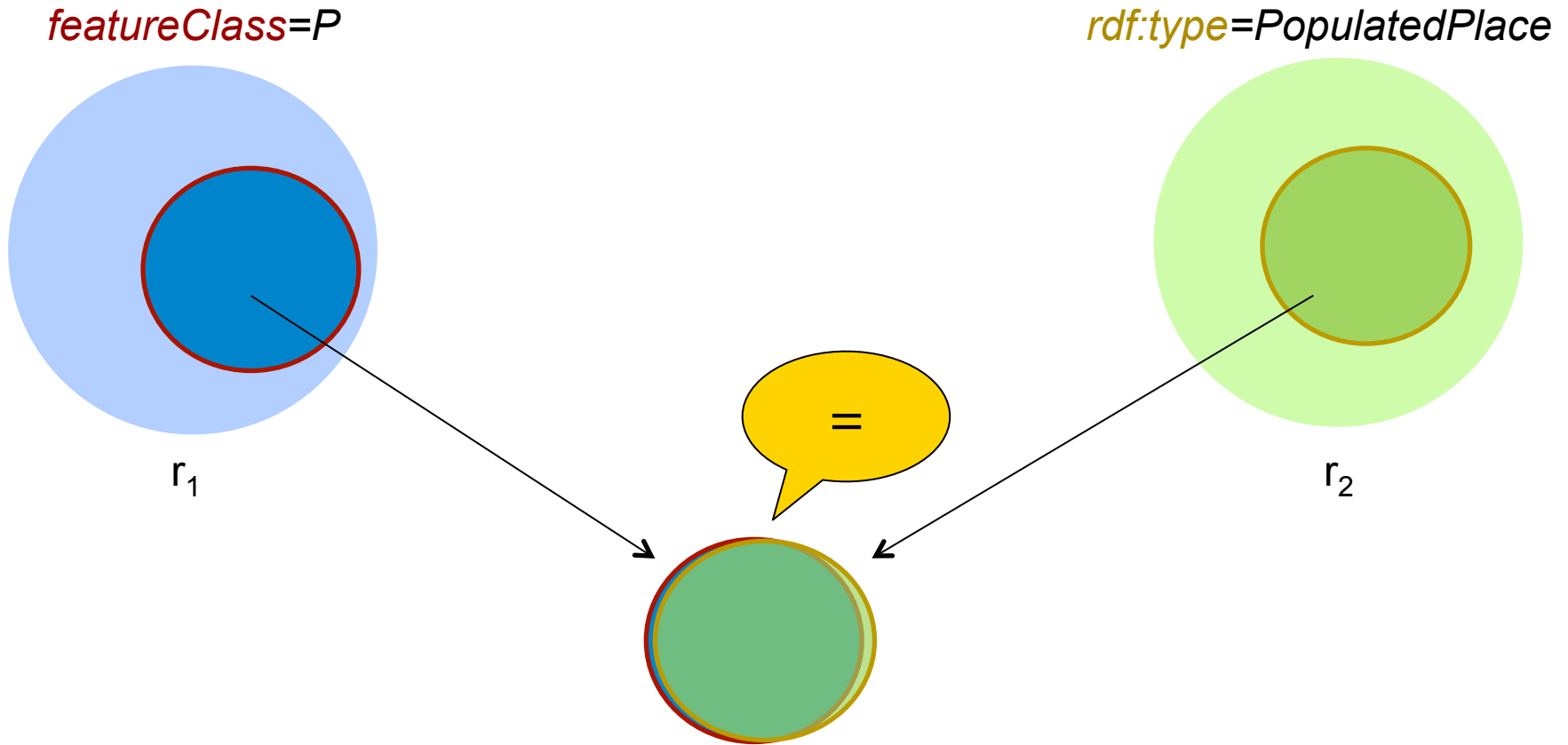
- Only a small number of Ontologies are linked
- Existing Concepts may not be sufficient for exhaustive set of alignments
- Alignments are necessary for the Interoperability goal of the Semantic Web

Solution: Generate Alignments using Linked Data – Extensional Approach

-  Represents set of instances belonging to ClassA
-  Represents set of instances belonging to ClassB



Aligning Restriction Classes Using Extensional Approach



$$\frac{|\text{Img}(r_1) \cap r_2|}{|\text{Img}(r_1)|} \approx 1 \quad \frac{|\text{Img}(r_1) \cap r_2|}{|r_2|} \approx 1$$

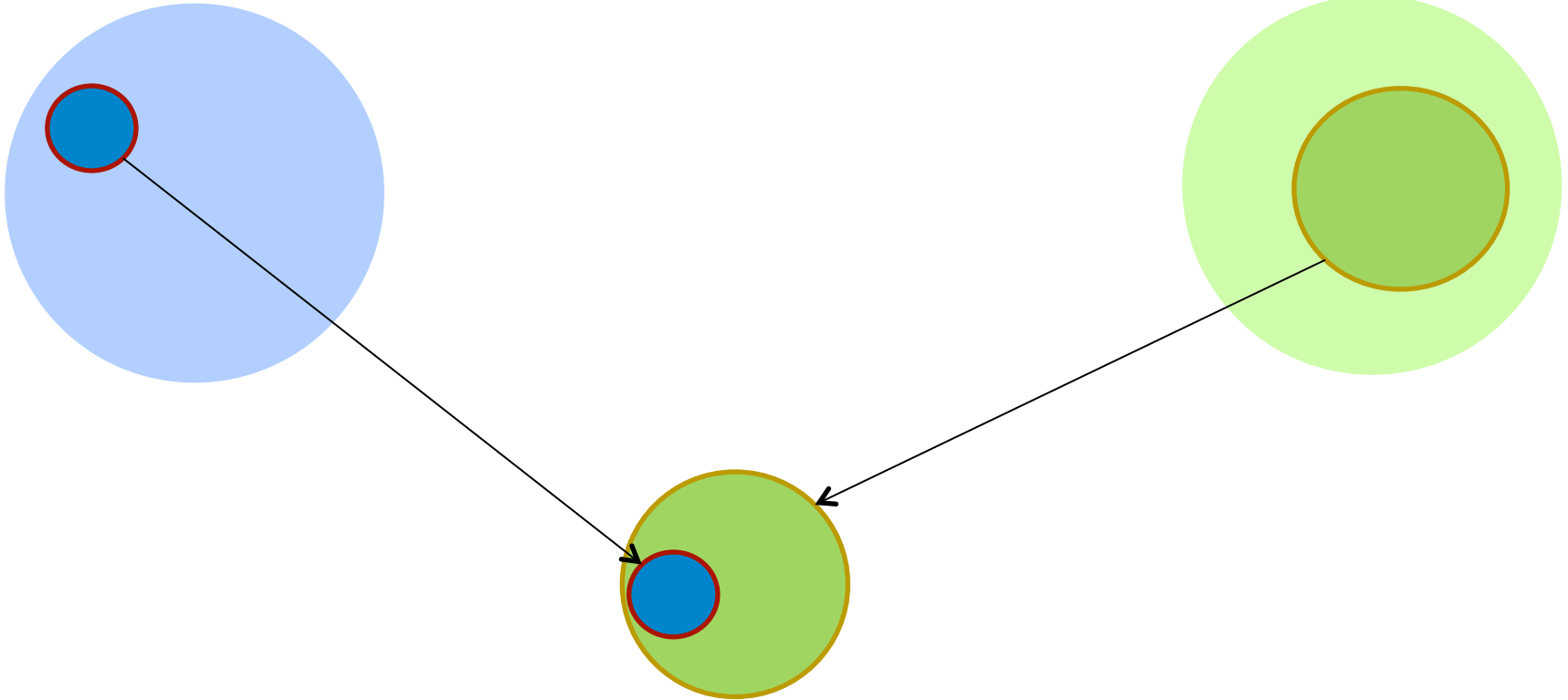
But there is a pattern to be explored in the subset relations

Let's look at 3 of the subset relations we found...

1) Schools in *GeoNames* are Educational Institutions in *DBpedia*

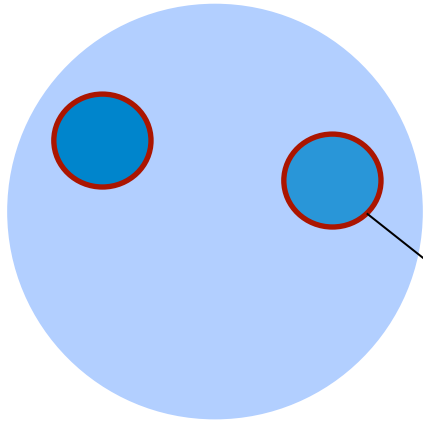
featureCode=S.SCH

rdf:type=EducationalInstitution



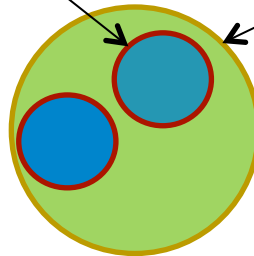
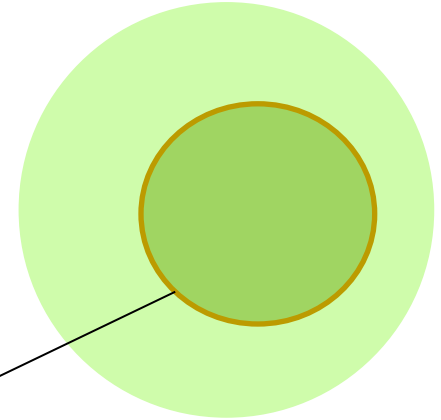
2) Colleges in *GeoNames* are Educational Institutions in *DBpedia*

featureCode=S.SCH



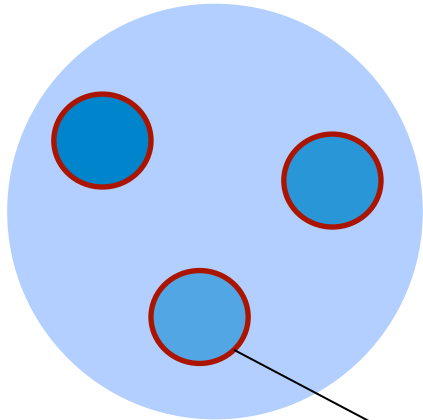
featureCode=S.SCHC

rdf:type=EducationalInstitution



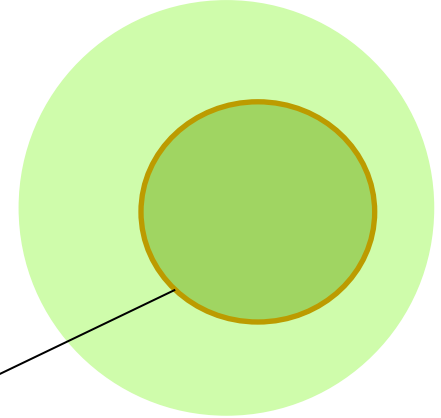
3) Universities in *GeoNames* are Educational Institutions in *DBpedia*

featureCode=S.SCH

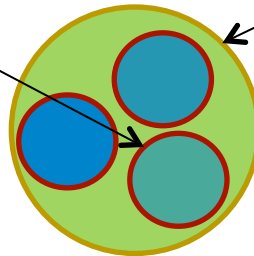


featureCode=S.SCHC

rdf:type=EducationalInstitution



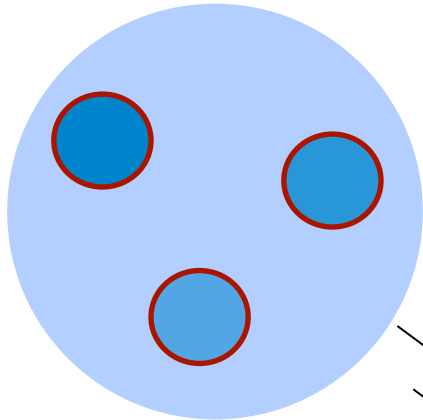
featureCode=S.UNIV



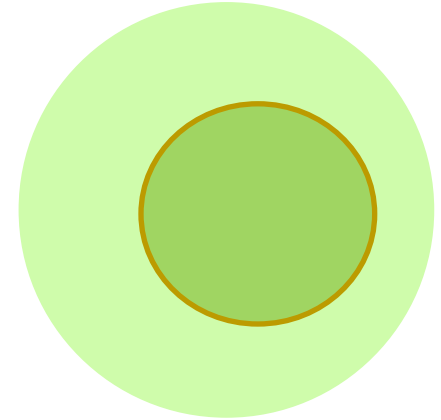
Taken by themselves, the subset relations are not useful

featureCode=S.SCH

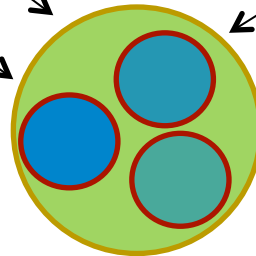
rdf:type=EducationalInstitution



featureCode=S.SCHC

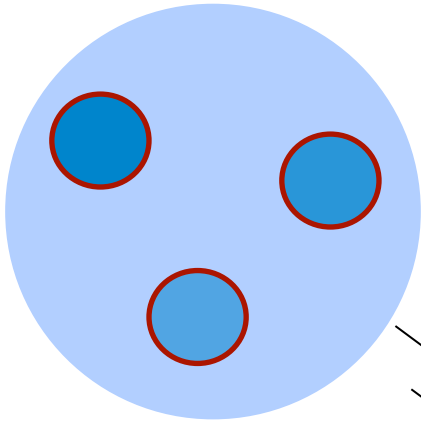


featureCode=S.UNIV



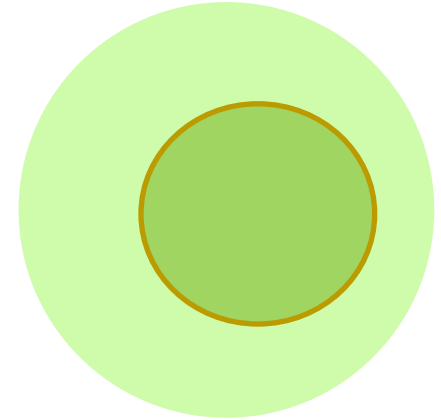
Using *featureCode* property as a hint, we form a *Union* of concepts

featureCode=S.SCH

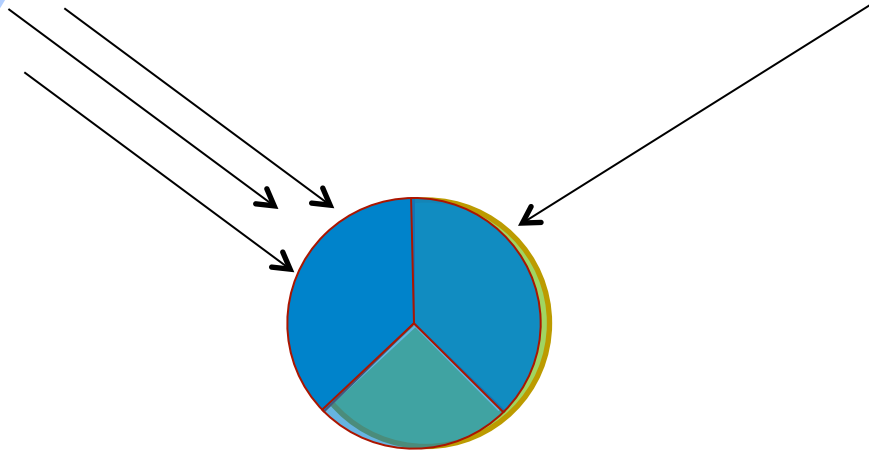


featureCode=S.SCHC

rdf:type=EducationalInstitution



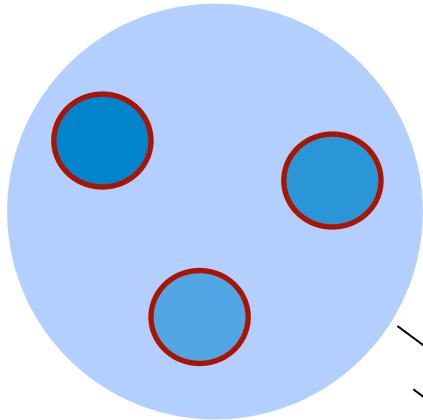
featureCode=S.UNIV



featureCode=S.SCH \cup *featureCode*=S.SCHC \cup *featureCode*=S.UNIV

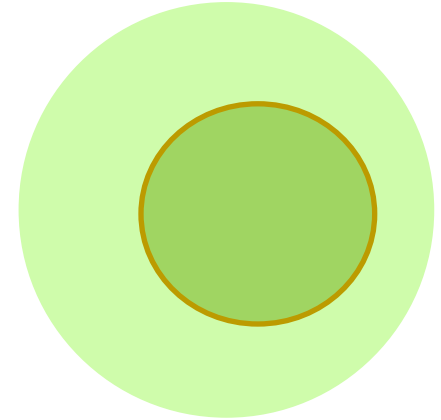
We Can Find Concept Coverings by Extensional Comparison (**Contribution 1**)

featureCode=S.SCH

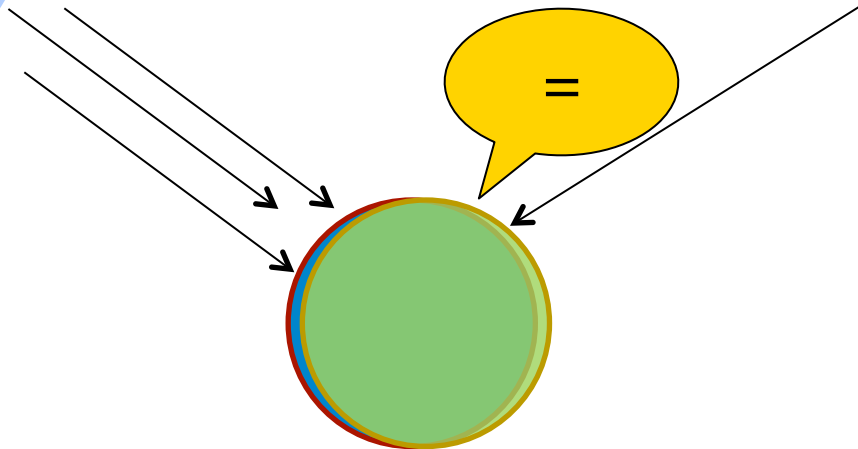


featureCode=S.SCHC

rdf:type=EducationalInstitution




featureCode=S.UNIV



featureCode=S.SCH \cup *featureCode*=S.SCHC \cup *featureCode*=S.UNIV

Example Alignments from DBpedia alignment with Geonames:

Larger Concept	Concepts Covered	Support	Outliers
<i>rdf:type</i> = dbpedia:Educational Institution	<i>geonames:featureCode</i> = {S.SCH, S.SCHC, S.UNIV}	396 out of 404 ($R'_U=0.98$)	S.BLDG (3/122), S.EST (1/13), S.LIBR (1/7), S.HSP (1/31), S.MUS (1/43)
<i>dbpedia:country</i> = dbpedia:Spain	<i>geonames:countryCode</i> = {ES}	3917 out of 3918 ($R'_U=0.99$)	IT (1/7635) 

Contribution 2: Finding Outliers

We found 7096 Alignments including Outliers for 5 pairs of sources aligned in Geospatial, Zoology & Genetics domains from 77966 subset relations

{parundek, knoblock}@usc.edu, ambite@isi.edu

**QUESTIONS? SEND US AN
EMAIL**

