

Leveraging Mathematical Subject Information to Enhance Bibliometric Data

Maria Koutraki, Olaf Teschke, Harald Sack, Fabian Müller, and Adam Bannister

28 May 2017, Portoroz Slovenia
Scientometrics@eswc2017

Agenda

- Introduction
 - Challenges in Mathematics
 - MSC classification system
- Approach
 - zbMATH data
 - Analysis
- Findings
- Conclusions

Why mathematics is challenging for bibliometrics?

1. Few publications and references → metrics are sensitive to distortions
 2. The unusual longevity of mathematical research
 3. The diverse nature of mathematics
- Quantitative methods like bibliometric analysis are prone to misrepresentations!

Need to consider subject information !!

Mathematical Subject Classification (MSC)

- Mathematical Subject Classification (MSC)
 - Classification scheme maintained by Mathematical Reviews, zbMATH
 - Introduced in 1970
 - Revised every 10 years
 - Current version: MSC2010
 - Linked to DBpedia
 - Three-level classification tree
 - 63 first-level nodes
 - >400 second-level nodes
 - >5000 leaf nodes

08-XX

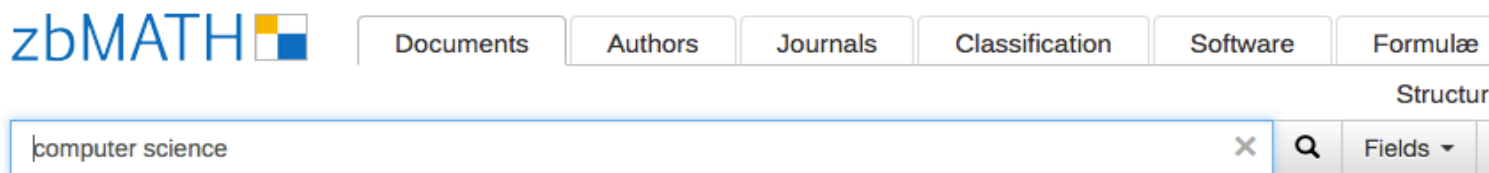
08Bxx

	General algebraic systems
	Varieties [See also 03C05]
08B05	Equational logic, Mal'cev (Mal'tsev) conditions
08B10	Congruence modularity, congruence distributivity
08B15	Lattices of varieties
08B20	Free algebras
08B25	Products, amalgamated products, and other kinds of
08B26	Subdirect products and subdirect irreducibility
08B30	Injectives, projectives

zbMATH Data

- Zentralblatt MATH (**zbMATH**) [1]: service for abstracting and reviewing pure and applied mathematical documents.
- Contains >3.5 M bibliographic entries.
- Coverage from 18th century up to now!
- Available online

[1]: <https://zbmath.org/>



The image shows the zbMATH website search interface. At the top left is the zbMATH logo. To its right are navigation buttons for Documents, Authors, Journals, Classification, Software, and Formulae. Below these is a search bar containing the text "computer science". To the right of the search bar are a search icon, a close icon, and a "Fields" dropdown menu.

Free access is limited to 3 results, and filter functions are disabled. For full access subscription is required.

Found 145631 documents (Results 1–100)

Zimmermann, Wolf; Picht, Roswitha

(Konzepte höherer Programmiersprachen. (to appear).) (German) Zbl 05834957

München: de Gruyter/Oldenbourg (ISBN 978-3-486-58102-7). 350 p. (2022).

MSC: 68N15 68N01 68-01

Filter recent res

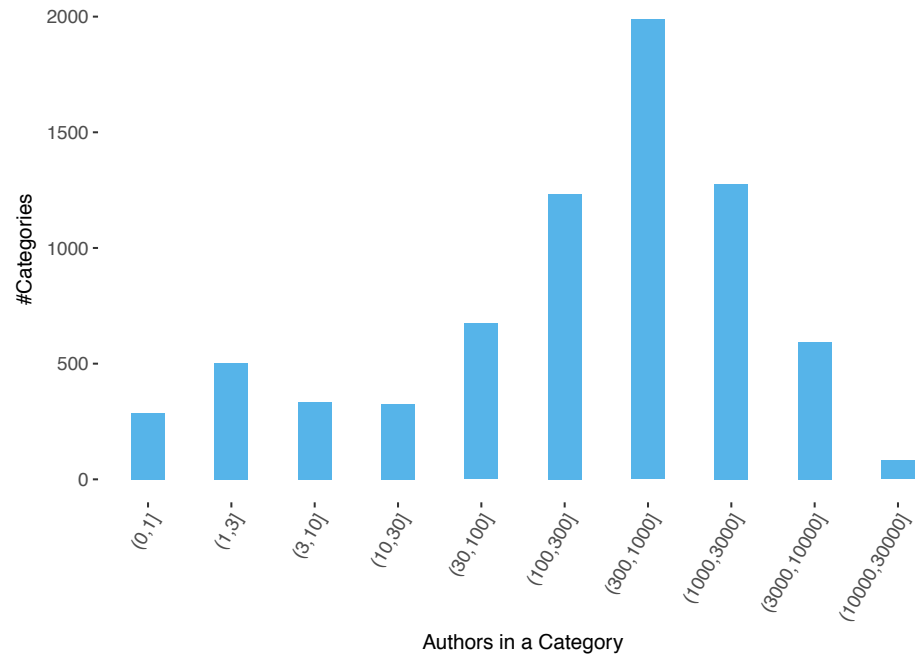
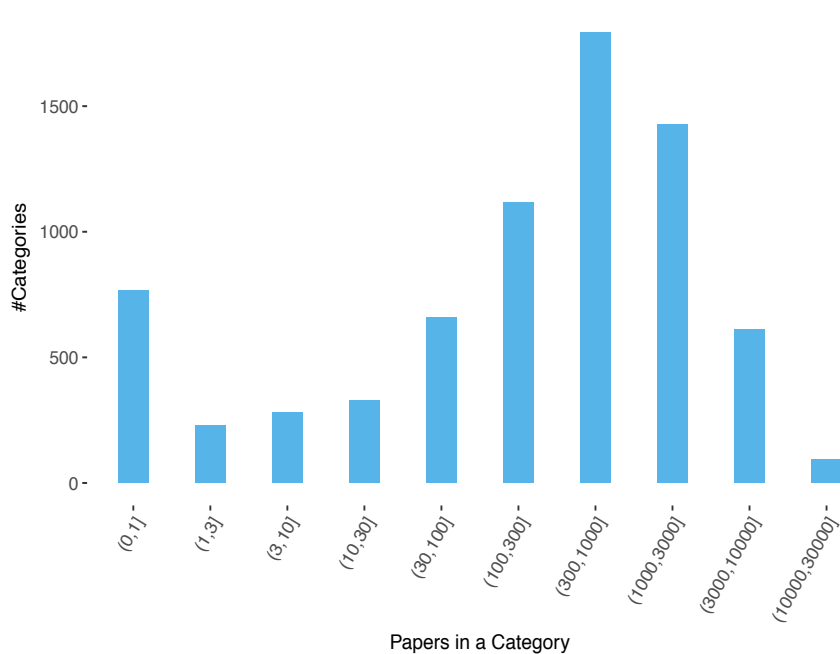
Authors

Fomin, Fedor V. (17

Saurabh Saket (10

Statistical Analysis of zbMATH Data (I)

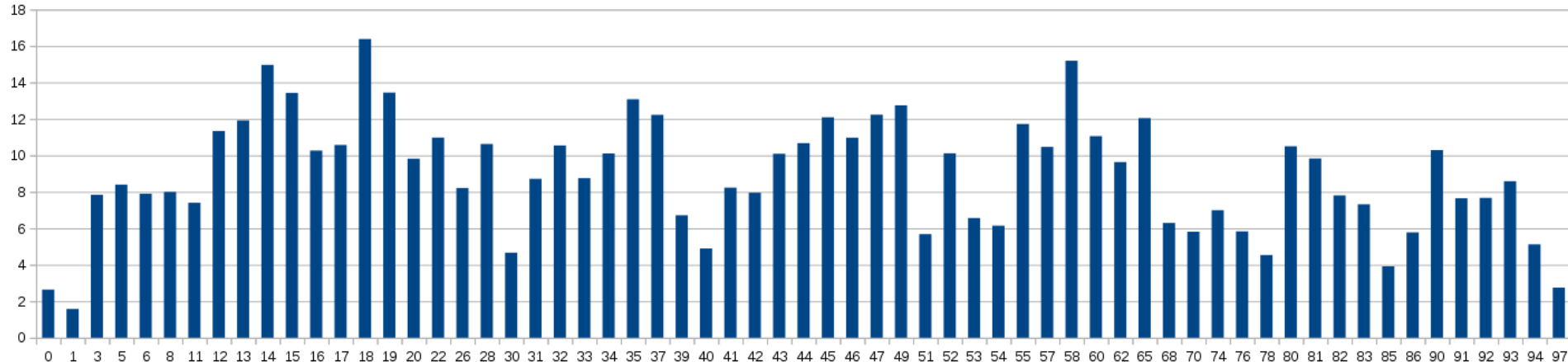
Distribution of zbMATH papers and authors in the different MCS categories.



Statistical Analysis of zbMATH Data (II)

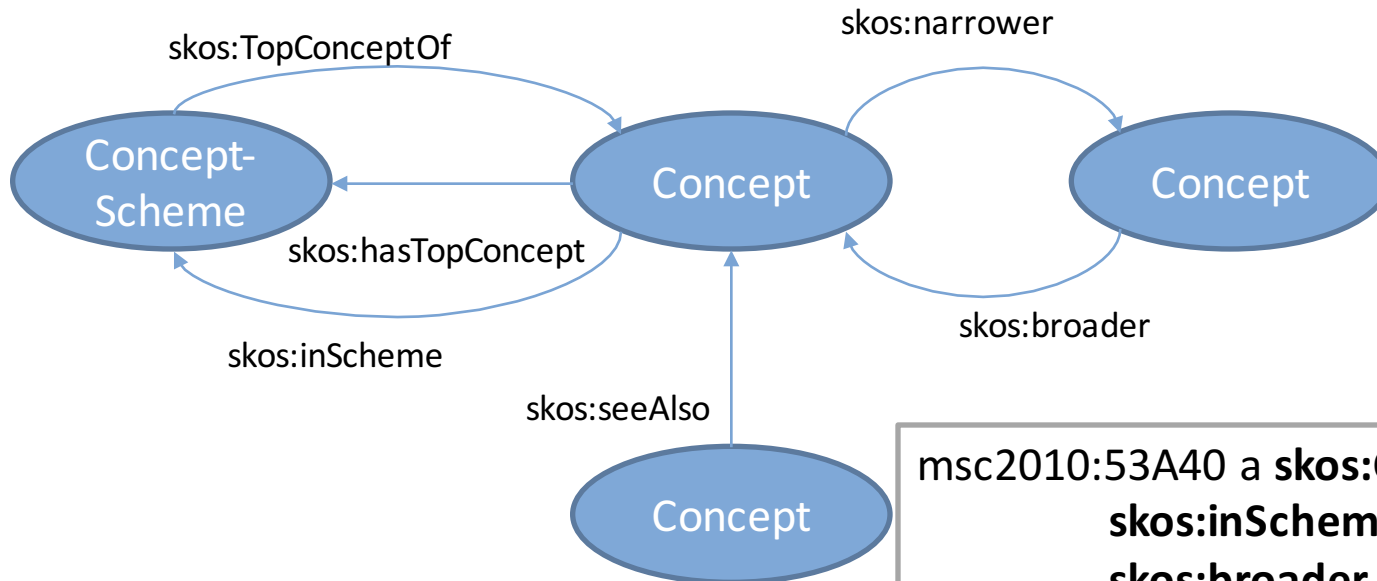
Citation frequency of top level MSC categories.

Highly diverse citation frequency for the top level MSC categories.



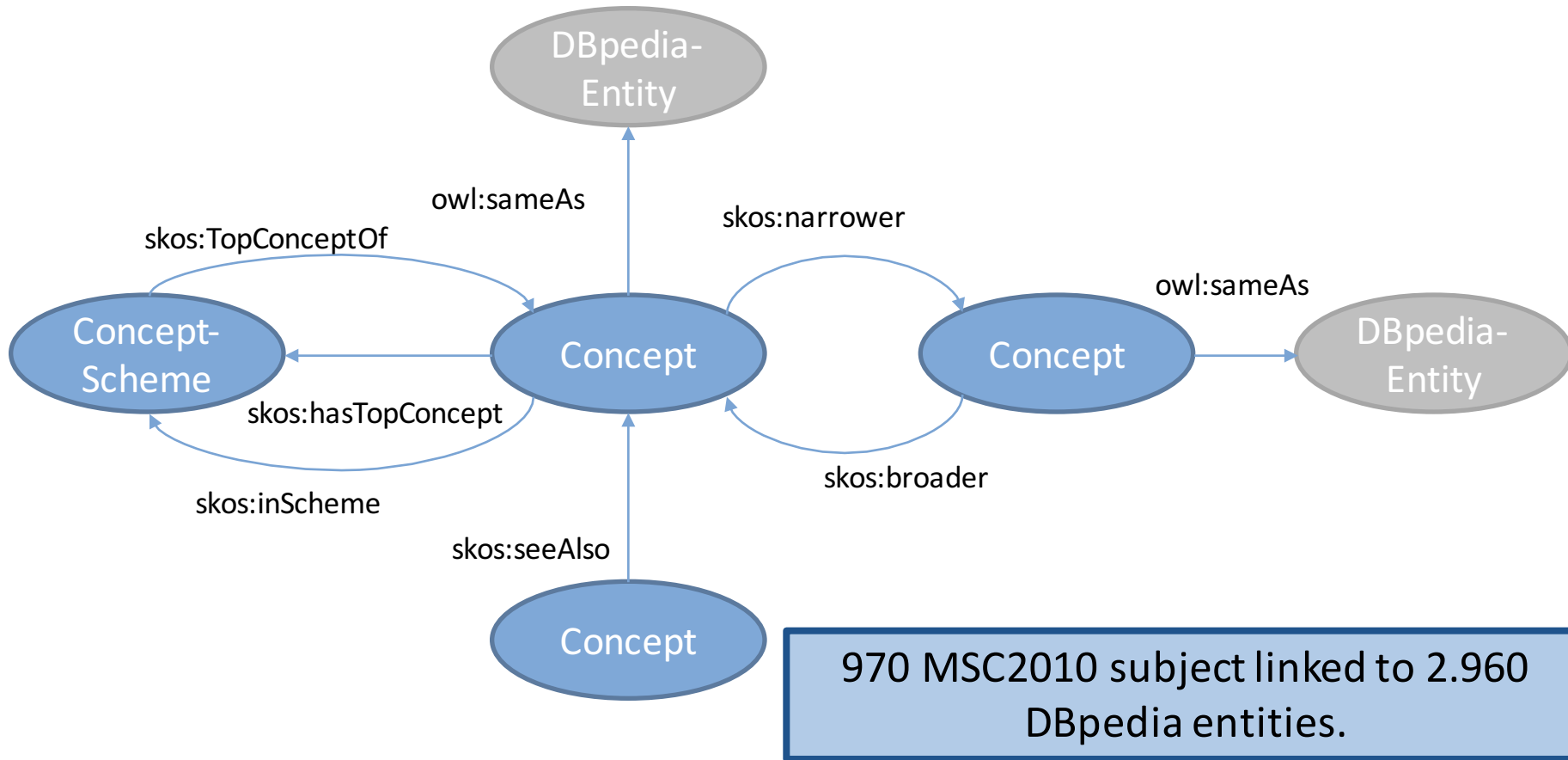
→ Necessity to consider subject information for bibliometric studies in mathematics.

Basic Hierarchy of MSC2010 (SKOS Core)



```
msc2010:53A40 a skos:Concept ;  
  skos:inScheme msc:2010: ;  
  skos:broader msc:53Axx ;  
  skos:preferredLabel "Other special  
  differential geometries@en" ;
```

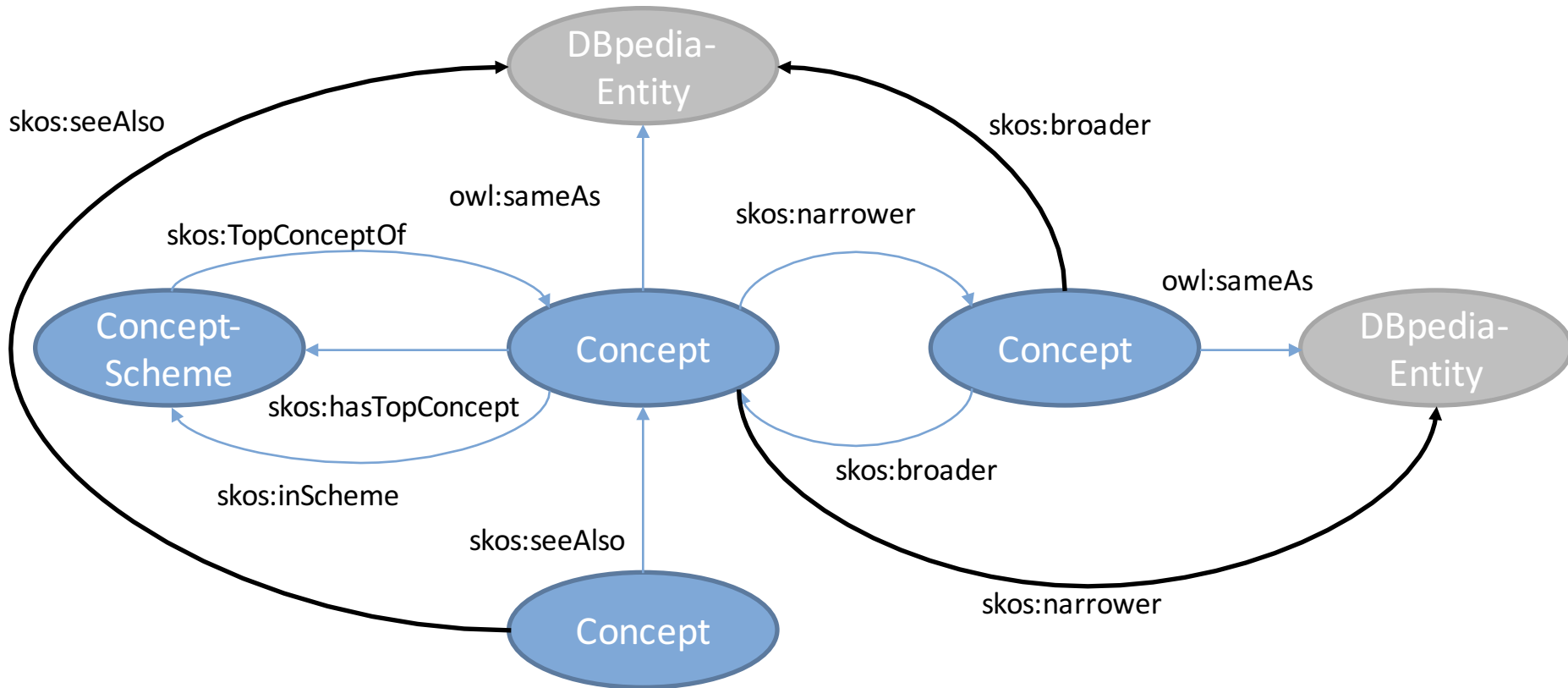

MSC2010 links to DBpedia [Lange et al. CICM'12]



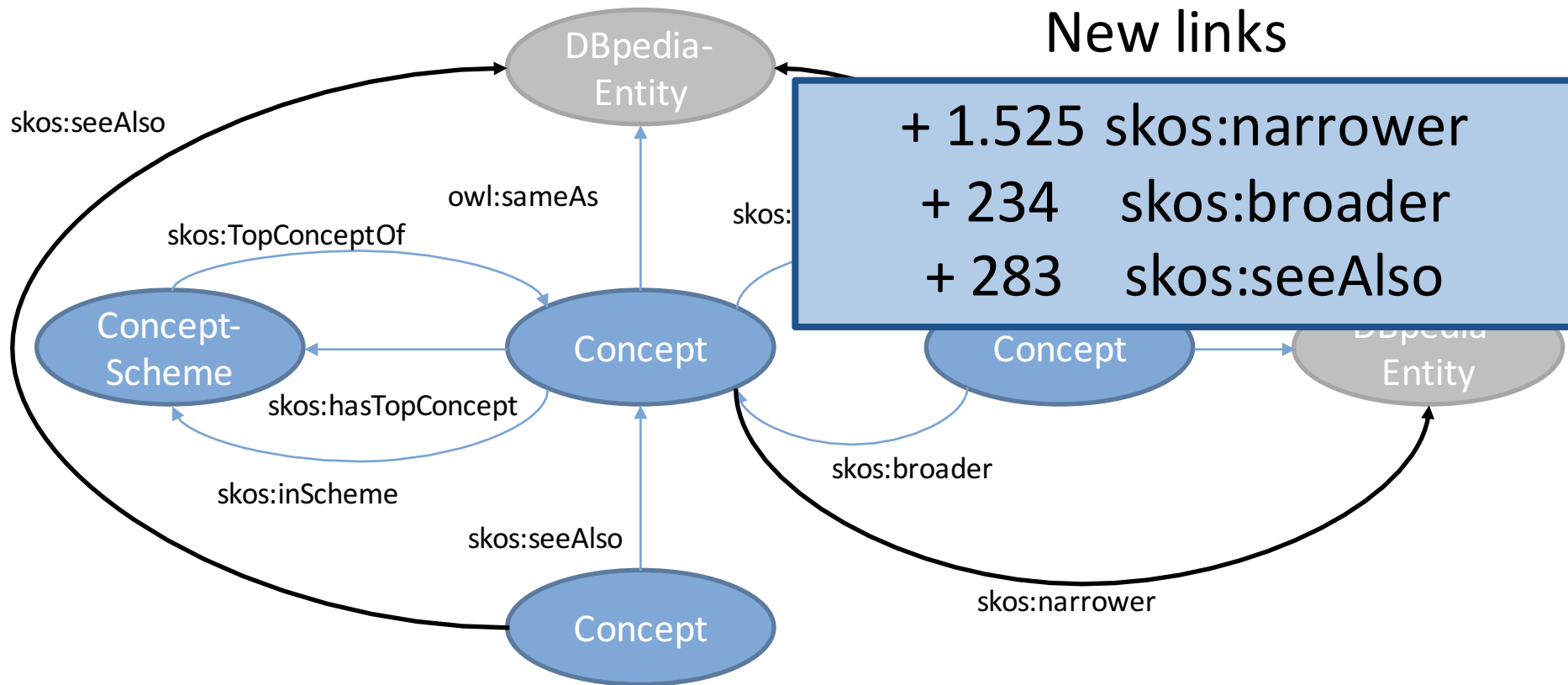
Adjustments on links to DBpedia

- Replace **redirects** with the URI of the original DBpedia entity.
- Replace the URIs of DBpedia entities **changed** from previous versions .
- Extension of links to DBpedia.

MSC2010 links to DBpedia -- Extended



MSC2010 links to DBpedia -- Extended



Similarity of MSC2010 Categories

- **Semantic similarity:** wr.t. DBpedia entities corresponding to MSC2010 categories.
 - Use similarity method proposed in [Piao et al.]
 - Properties of the resources
 - Other similarity measures: equal self-similarity, symmetry, minimality
- **Statistical similarity:** w.r.t MSC2010 categories assigned to zbMATH paper collection.

- Symmetric

$$Jaccard(cat_a, cat_b) = \frac{cat_a \cap cat_b}{cat_a \cup cat_b}$$

- Asymmetric

$$Jaccard(cat_a \rightarrow cat_b) = \frac{cat_a \cap cat_b}{cat_a}$$

Experimental Setup - Results

- **Statistical similarity:** zbMATH collection of papers a.k.a. 100K papers.
- **Semantic similarity:** Top 10K category pairs from statistical similarity.
 - Symmetric: 937 mapped to DBpedia via owl : sameAs .
 - Asymmetric: 672 mapped to DBpedia via owl : sameAs .
- Correlations coefficient of semantic and statistical similarities

Similarity measures	r
Semantic – Jaccard	0.34
Semantic – asymmetric Jaccard	-0.04

Issues in MSC2010 (1/2)

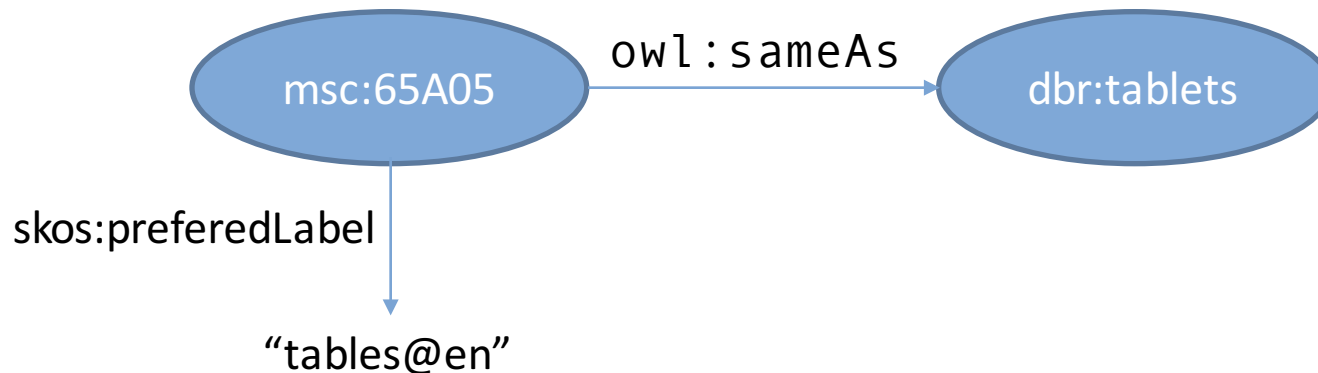
- Structural Issues

14-XX		Algebraic geometry
	14Txx	Tropical geometry [See also 12K10, 14M25, 14N10, 52B20]
	14T05	Tropical geometry [See also 12K10, 14M25, 14N10, 52B20]
	14T99	None of the above, but in this section

Super-category	Sub-category	Label
13Gxx	13G05	Integral domains
14Txx	14T05	Tropical geometry
22Cxx	22C05	Compact groups
45Qxx	45Q05	Inverse problems
62Qxx	62Q05	Statistical tables
65Axx	65A05	Tables
85-XX	85Axx	Astronomy and astrophysics
...

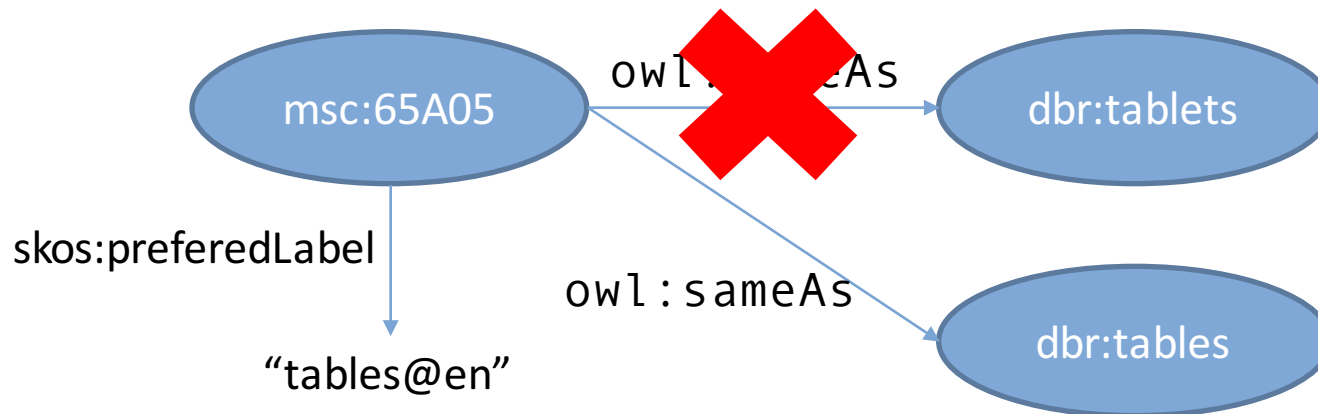
Issues in MSC2010 (2/2)

- Imprecise owl:sameAs links
 - Links to **redirects** resources to DBpedia.
 - **Erroneous** links between the MSC2010 and DBpedia.
 - E.g. :



Issues in MSC2010 (2/2)

- Imprecise owl:sameAs links
 - Links to **redirects** resources to DBpedia.
 - **Erroneous** links between the MSC2010 and DBpedia.
 - E.g. :



Conclusions

- Correct, upgrade, and extend existing MSC2010 mappings to DBpedia.
- Computation of MSC categories similarity through statistical and semantic measures.
- Detection of inconsistencies in MSC2010.

Questions?

Thank you !

We are hiring!
Come and talk to us!
ISE group @AIFB KIT

Page



These documents are intended for presentation purposes only.
Copyright lies with FIZ Karlsruhe.
Any distribution or use of these documents or part thereof is
subject to FIZ Karlsruhe's express approval.

© FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur GmbH 2017