gesis
Leibniz Institute for the Social Sciences

New Ways of Mapping Knowledge Organization Systems

Using a Semi-Automatic Matching-Procedure for Building Up Vocabulary Crosswalks

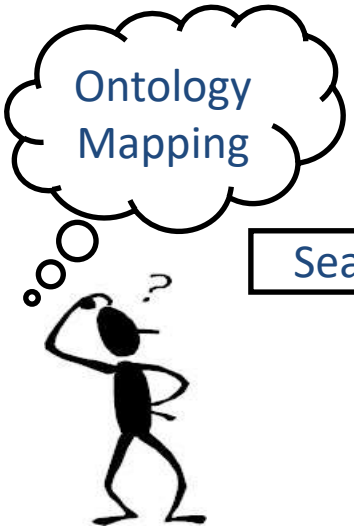
Andreas Oskar Kempf – GESIS – Leibniz Institute for the Social Sciences
Benjamin Zapilko – GESIS – Leibniz Institute for the Social Sciences
Dominique Ritze – Mannheim University Library
Kai Eckert – Mannheim University

Member of the Leibniz Association

Content

- Vocabulary Crosswalks
  - Why are they needed?
  - How do they look like?
- Automatic Matching Initiatives and Procedures
  - Ontology Matching Approaches
- OAEI Library Track 2012
  - What kind of outcome and limitations regarding an automatic creation of vocabulary crosswalks do we have to expect?
- Optimizing the Manual Evaluation Process
- Conclusion and Outlook

Mapping KOS - Motivation

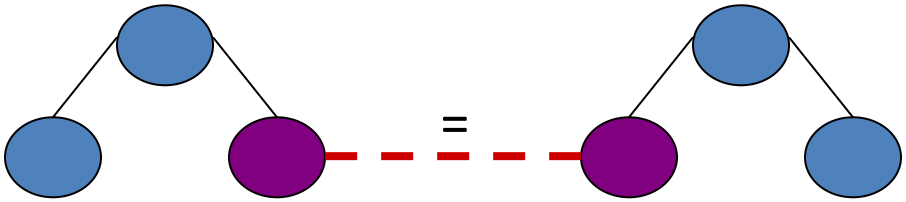


Search

0 results

Publication x
subject (thesaurus 2): ontology alignment

Thesaurus 1 Thesaurus 2



Ontology Mapping Ontology Alignment

Search

Publication x
subject (thesaurus 1): ontology alignment

Member of the Leibniz Association

1. Vocabulary Crosswalks (1/2)

Why are they needed?

- allow for integrated and high-quality search scenarios in distributed information collections indexed on the basis of different controlled vocabularies
- allow for interoperability among different knowledge organization systems
- allow for vocabulary expansion and provide possible routes into various domain-specific languages
- allow for query expansion and reformulation
- allow for the use of familiar vocabularies to maneuver between different information resources

1. Vocabulary Crosswalks (2/2)

How do they look like?

- consist of equivalence (=), hierarchy (</>) and association (^) relations
- could consist of a mapping to several terms of the vocabulary being mapped to and of a combination of terms of the vocabulary being mapped to
- are established bilaterally (A > B and B > A)

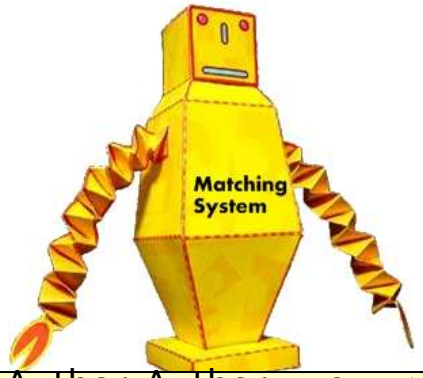
How are they being done?

- get an overview over the topical overlap and the structure of the different vocabularies
- build up an understanding of the meaning and semantics of the terms and the internal relations of the vocabularies
- start the mapping process (take all the internal relations, synonyms/non-descriptors within the concepts into account)
- modify mappings already built up during the mapping process
- perform retrieval tests

Projects

- MACS (National Libraries CH, F, GB, GER), OCLC Mappings

Ontology Matching



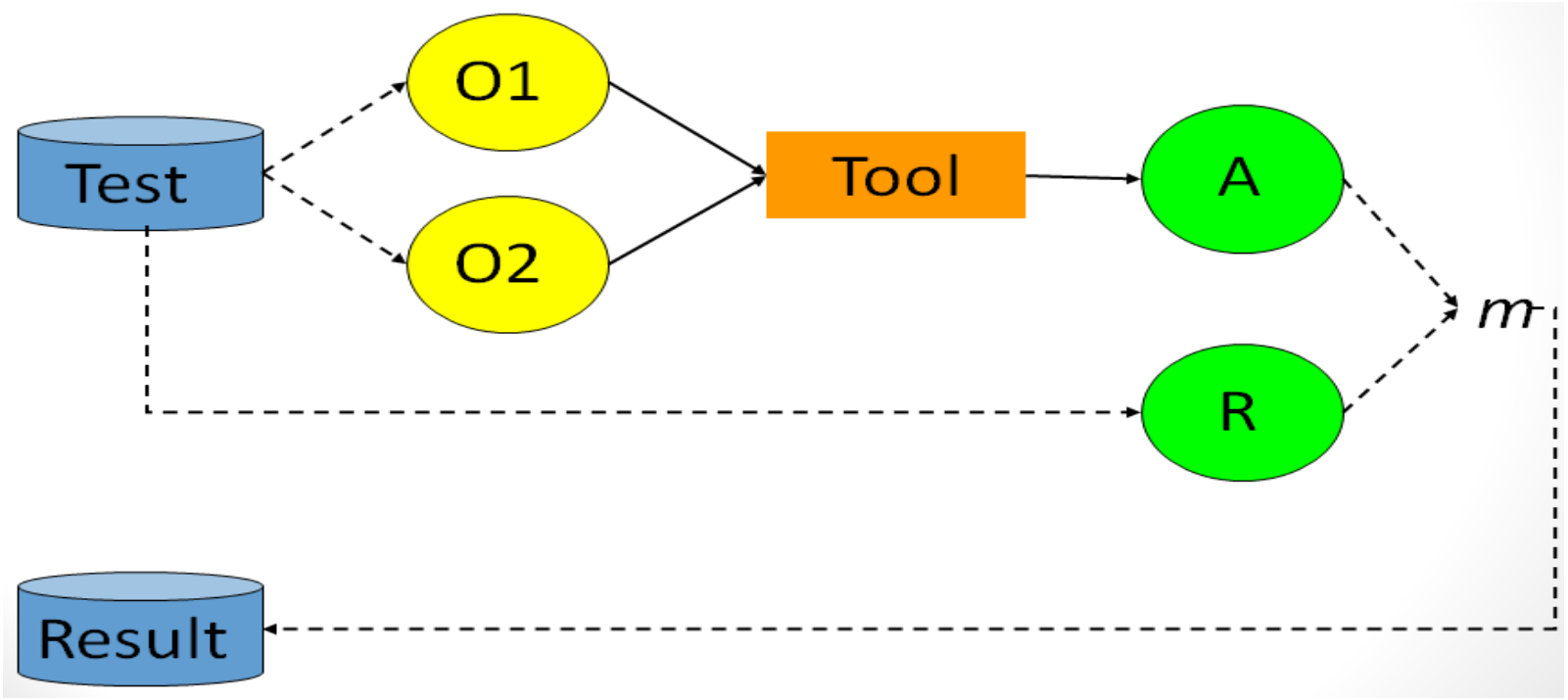
Person
Author
CommitteeMember
PCMember
Document
reviews
Paper
writes
Review

< Author, Author, =, 0.97 >
< Paper, Paper, =, 0.94 >
< reviews, reviews, =, 0.91 >
< writes, writes, =, 0.7 >
< Person, People, =, 0.8 >
< Document, Doc, =, 0.7 >
< Reviewer, Review, =, 0.6 >
...

People
Author
Reviewer
writes
reviews
Doc
Paper

Member of the Leibniz Association

Ontology Matching Evaluation



2. Terminology Mapping (2/2)

Ontology Alignment Evaluation Initiative (OAEI)

- Annual international campaign started in the year 2004
- Different tracks/datasets
- Objectives:
  - Improving the performances of mapping tools in the field of ontology matching
  - Comparing the different algorithms
  - Detecting new challanges for matching systems

Member of the Leibniz Association

OAEI Library Track 2012

Library

The library track is a real-word task to match the STW and the TheSoz social science thesauri in SKOS. The goal of this track is to find whether the matchers can handle these lightweight ontologies including a huge amount of concepts and additional descriptions. Results will be avaluated both against a reference alignment and through manual scrutiny of alignments.

Data Sets

- Thesaurus for the Sociel Sciences (TheSoz)
  about 8.000 concepts with about 4.000 additional keywords/entry terms (EN, DE, FR)
- Thesaurus for Economics (STW)
  about 6.000 concepts with about 19.000 additional keywords/entry terms (EN, DE)

gesis Leibniz Institute for the Social Sciences

ZBW Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Reference Alignment (2006)

- TheSoz > STW; STW > TheSoz
  (≈7,000 intellectually created relations in each direction)

Member of the Leibniz Association

Thesaurus = Ontology?

Thesauri: Polydimensional Ontologies (for they are characterized by only a limited number of conceptual relation types).

Ontologies: Multidimensional Systems with potentially infinite number of relation types.

See: Gietz 2001: 24f.

SKOS	OWL
skos:Concept	owl:Class
skos:prefLabel skos:altLabel	rdfs:label
skos:scopeNote skos:notation	rdfs:comment
A skos:narrower B	A rdfs: subClassOf B
A skos:broader B	B rdfs:subClassOf A
skos:related	rdfs:seeAlso

Tropical Fruit

Ananas

Metal Product -> Metal

Results

System	Precision	Recall	F-Measure	Time (s)	Size	1:1
GOMMA	0.537	0.906	0.674	804	4712	
ServOMapLt	0.654	0.687	0.670	45	2938	
LogMap	0.688	0.644	0.665	95	2620	
ServOMap	0.717	0.619	0.665	44	2413	yes
YAM++	0.595	0.750	0.664	496	3522	
LogMapLt	0.577	0.776	0.662	21	3756	
G02A	0.675	0.645	0.660	32773	2671	
Hertuda	0.465	0.925	0.619	14363	5559	
WeSeE	0.612	0.607	0.609	144070	2774	yes
HotMatch	0.645	0.575	0.608	14494	2494	yes
CODI	0.434	0.481	0.456	39869	3100	yes
MapSSS	0.520	0.184	0.272	2171	989	yes
AROMA	0.107	0.652	0.184	1096	17001	
Optima	0.321	0.072	0.117	37457	624	

Manual Evaluation

	Equivalence Relations (in total)	**Correct Equivalence Relations**	Non-Correct Equivalence Relations
AROMA	3.500	**215 (6,1%)**	3.285
CODI	628	**162 (25,8%)**	466
GO2A	631	**213 (33,8%)**	418
GOMMA	682	**246 (36,1%)**	436
Hertuda	828	**269 (32,5%)**	556
HotMatch	448	**194 (43,3%)**	254
LogMapLt	540	**234 (43,3%)**	306
LogMap	403	**203 (50,4%)**	200
MapSSS	175	**64 (36,6%)**	111
Optima	165	**38 (23,0%)**	127
ServOMapL	525	**252 (48,0%)**	273
ServOMap	433	**232 (53,8)**	201
WeSeE	682	**225 (33,0%)**	457
YAM++	613	**248 (40,5%)**	365

Optimizing the Evaluation Process

Leading question:

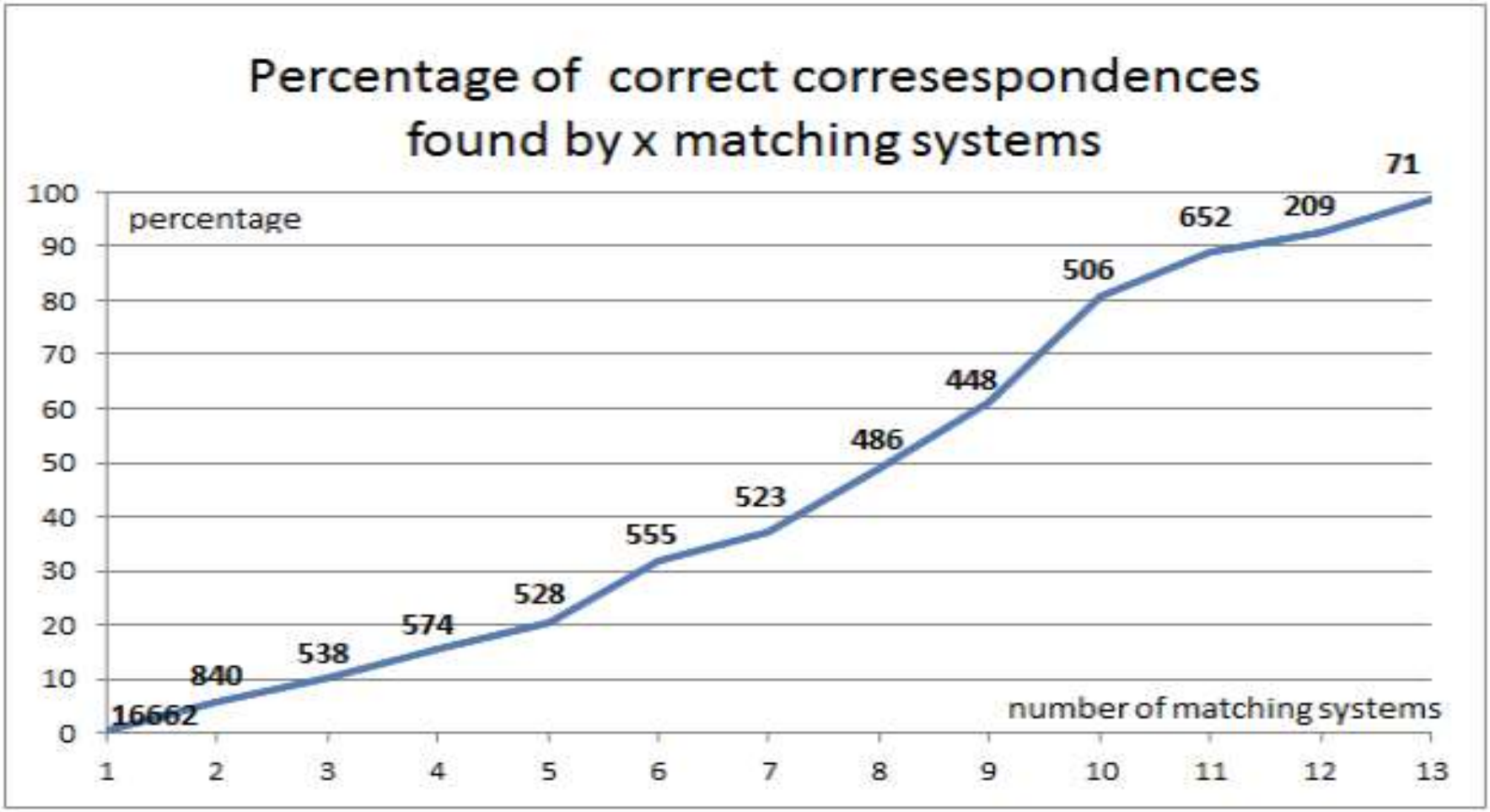
How can the intellectual matching process be best supported by ontology matching tools?

Approach:

Reorganizing the alignments according to the largest agreement between the different matching tools.

	All correspondences (including duplicates)	Unique correspondences
Total number	55466	22592
…of which are correct	21541	2484 (11%)

Number of Accordances between the different Matching Tools



Percentage of Correct Correspondences

Number of corresponding matchers	Number of all correspondences	Number of all correct correspondences	Percentage of correct correspondences
13	71	70	98.56 %
12	209	194	92.82 %
11	652	581	89.11 %
10	506	409	80.83 %
9	448	275	61.38 %
8	486	238	48.87 %
7	523	194	37.09 %
6	555	177	31.89 %
5	528	108	20.45 %
4	574	90	15.68 %
3	538	56	10.41 %
2	840	48	5.71 %
1	16662	50	0.27 %

Comparison between Regular and Optimized Evaluation Scenario

			optimized scenario	*optimized scenario*	*normal evaluation*	*normal evaluation*
No. of corresponding matchers	No. of all correspondences	% of all correspondences (22592=100%)	No. of correct correspondences	% of all correct correspondences (2484=100%)	No. of correct correspondences (estimated)	% of all correct correspondences (2484=100%)
13	71	0.31 %	70	2.82 %	8	0.32 %
12	280 (71 + 209)	1.24 %	264	10.63 %	31	1.25 %
11	932 (…+…)	4.13 %	845	34.02 %	103	4.15 %
10	1438 (…+…)	6.37 %	1254	50.48 %	158	6.36 %
9	1886 (…+…)	8.34 %	1529	61.55 %	207	8.33 %
8	2372 (…+…)	10.50 %	1767	71.14 %	261	10.51 %
7	2895 (…+…)	12.81 %	1961	78.95 %	318	12.80 %
6	3450 (…+…)	15.27 %	2138	86.1 %	380	15.30 %
5	3978 (…+…)	17.61 %	2246	90.42 %	438	17.63 %
4	4552 (…+…)	20.15 %	2336	94.04 %	501	20.17 %
3	5090 (…+…)	22.53 %	2392	96.30 %	560	22.54 %
2	5930 (…+…)	26.25 %	2440	98.23 %	652	26.25 %
1	22592 (…+…)	100 %	2484	100 %	2484	100 %

Conclusion

- Significant differences between the different ontology matching tools
- Some tools provide rather promising performances
- None of the evaluated matching tools alone could ensure high-quality standards for building up vocabulary crosswalks automatically
- Ontology matching tools can be used to optimize the intellectual evaluation process
- By reorganizing the validation process considering the number of accordances between the different matching tools the intellectual evaluation process could be made more time-efficient
- Matching tools can be used as recommendation systems for manual evaluation

Thank you for your attention.

Contact

Dr. Andreas Oskar Kempf
GESIS – Leibniz-Institute for the Social Sciences
andreas.kempf@gesis.org

Benjamin Zapilko
GESIS – Leibniz-Institute for the Social Sciences
benjamin.zapilko@gesis.org

Dominique Ritze
Mannheim University Library
dominique.ritze@bib.uni-mannheim.de

Kai Eckert
Mannheim University
kai@informatik.uni-mannheim.de

Member of the Leibniz Association