

Ontology matching

Cássia Trojahn dos Santos
UTM & IRIT

3rd Cameleon Projet Workshop - 20 décembre 2012

Agenda

1. Background : semantic web, linked data, ontologies
2. Ontology matching
3. Ontology matching evaluation
4. Multilingual ontology matching

(Traditional) Web for humans

Fnac. On ne peut qu'adhérer

KOBO & EBOOKS | FNACPRO | ARGUS REPRISE | MY FNAC | ÉVÉNEMENTS | COMMUNAUTÉS | SAV

High-tech

Mon Compte Aide

CADEAUX DE NOËL

Notre boutique de Noël

24h

Livr à temps

RETRAIT MAGASIN GRATUIT

LIVRA

Samsung Galaxy S3 Mini (i8190) - Blanc

Smartphone sous Android OS - Samsung

Donner votre avis

349€90

LIVRÉ À TEMPS POUR NOËL
La livraison Express à seulement 4,99 € Conditions

TELEPHONE (5) Payez en 5 fois soit 71,38 € par mois J'en profite

En Stock Livré à temps pour Noël avec la livraison Express *

22 neufs à partir de 329€ 1 d'occasion à partir de 305€

MAGASIN Retrait gratuit dans votre FNAC (?)

LIVRAISON Livraison à partir de 4,99€ (?) Livré à temps pour Noël avec la livraison express, encore 13h36 pour être livré vendredi 21 décembre (?)

LES PLUS FNAC Satisfaire ou rembourser pendant 15 jours Garantie Fnac 1 an inclus SAV Fnac, 400 experts à votre service SAV Fnac Reprise, la Fnac rachète votre ancien matériel

Concept Talle écran

Mono-bloc 4"

Web for machines

Home > Samsung phones > Samsung GALAXY Nexus CDMA

Samsung GALAXY Nexus CDMA

[Follow](#)

[Specs](#) [Rivals](#) [Reviews \(31\)](#) [Discuss \(2\)](#) [News \(133\)](#) [Price](#) [Size](#) [360°](#) [Video](#) [Pics](#)



[VIEW ALL PHOTOS \(31\)](#)

[COMPARE](#) [BOOKMARK](#)

Announced: 21 Oct 2011
Market status: Released
Release date: 15 Dec 2011

US CARRIER AVAILABILITY:

DESIGN

Device type:

OS:

Dimensions:

Weight:

DISPLAY

Physical size:

Resolution:

Pixel density:

Technology:

Colors:

Touchscreen:

Features:

BATTERY

Talk time:

Stand-by time:

Source of: http://www.phonearena.com/phones/Samsung-GALAXY-Nexus-CDMA

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
<title>Samsung GALAXY Nexus CDMA specs</title>
<meta property="fb:app_id" content="152451818130304" />
<meta property="og:site_name" content="Phone Arena" />
<meta property="og:type" content="article" />
<meta property="og:url" content="http://www.phonearena.com/phones/Samsung-GALAXY-Nexus-CDMA" />
<meta property="og:image" content="http://i-cdn.phonearena.com/images/phones/31943-specs/Samsu..."/>
<meta property="og:title" content="Samsung GALAXY Nexus CDMA specs" />
<meta property="og:description" content="The Samsung GALAXY Nexus for Verizon is the first Andro..."/>
<meta name="generator" content="Phone Arena" />
<meta name="keywords" content="" />
<meta name="description" content="The Samsung GALAXY Nexus for Verizon is the first Android I..."/>
<meta http-equiv="content-language" content="en_US" />
<meta name="verify-v1" content="yCKjtGDRRxyOZTeD1N082zeqhTNxD2zbMuUcz-TCo" />
<meta name="google-site-verification" content="ycCxjGDRRxyOZTeD1N082zeqhTNxD2zbMuUcz-TCo" />
<link rel="shortcut icon" href="http://s-cdn.phonearena.com/images/favicon.ico?r=1" type="image/x-icon" />
<link rel="alternate" type="application/rss+xml" title="Phone Arena - Latest News and Articles" href="http://www.phonearena.com/rss.xml" />
<link rel="alternate" type="application/rss+xml" title="Phone Arena - Latest Reviews" href="http://www.phonearena.com/reviews/rss.xml" />
<link rel="alternate" type="application/rss+xml" title="Phone Arena - Latest Phones" href="http://www.phonearena.com/phones/rss.xml" />
<link rel="stylesheet" href="http://s-cdn.phonearena.com/?b=css&f=screen.css,arena.css.widget" />
```

Line 2, Col 92

4 / 27

Web for machines

ebay [Browse by category](#)


Samsung Galaxy Nexus SCH-I515 - 32GB - Metallic silver (Verizon) Smartphone
User reviews (62) ★★★★★

Top pick  [Seller refurbished](#) [Free shipping](#)

Extracted structured data

rdfa-node	property:	value:
	image:	http://i.ebayimg.com/00/\$KGrHqR_lqE63Z2gurWBO7d7dTLow~~_32.JPG?set_id=89040003C1
	type:	ebay-objects:item
	uri:	http://www.ebay.com/ctg/Samsung-Galaxy-Nexus-SCH-I515-32GB-Metallic-silver-Verizon-Smartphone-i110653084
	title:	Samsung Galaxy Nexus SCH-I515 - 32GB - Metallic silver (Verizon) Smartphone
	site_name:	eBay
	app_id:	102628213125203

 Seller refurbished
 Top-rated seller

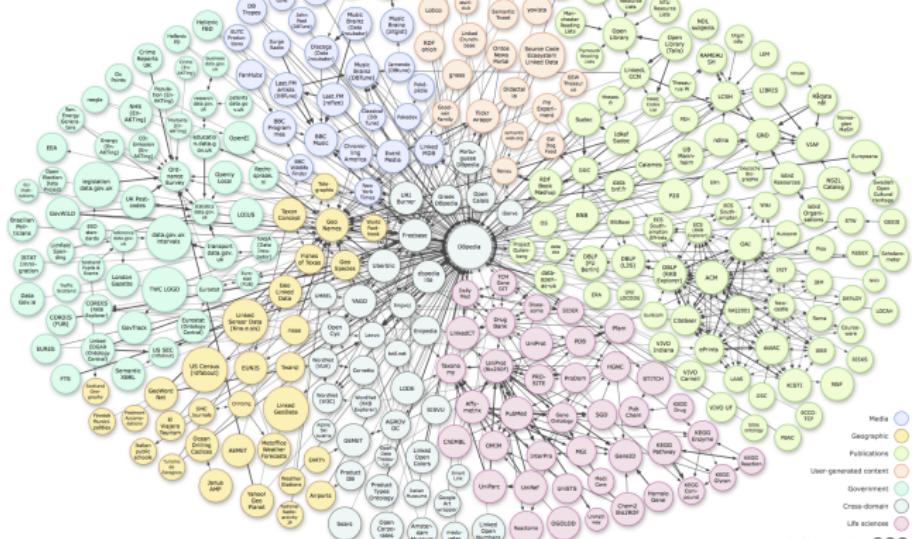
Buy it now [Buy it now](#) 23d 23h 51m

Semantic web

The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.

(“The Semantic Web”, Berners-Lee et al. 2001)

(Open) linked data



As of September 2011



Querying the open linked data

<http://dbpedia.org/sparql>

Virtuoso SPARQL Query Editor

About | Namespace Prefixes | Inference rules | SPARQL

Default Data Set Name (Graph IRI)
<http://dbpedia.org>

Query Text

```
SELECT * WHERE {
  <http://dbpedia.org/resource/Samsung_S8300>
  ?prop
  ?value .
}
```

http://xmlns.com/foaf/0.1/name	"Samsung S8300 Ultra Touch"@en
http://purl.org/dc/terms/subject	http://dbpedia.org/resource/Category:Samsung mobile phones
http://dbpedia.org/property/name	"Samsung S8300 Ultra Touch"@en
http://dbpedia.org/property/size	"762.0^^< http://dbpedia.org/datatype/second >
http://dbpedia.org/property/storage	"80^^^< http://www.w3.org/2001/XMLSchema#int >
http://dbpedia.org/property/battery	"Li-Ion"@en
http://dbpedia.org/ontology/abstract	The Samsung S8300 (also known as Ultra Touch or Tocco Ultra) is a high-spec mobile phone model from Samsung Telecommunications. Released over April 2009. Renowned for its AMOLED Touch-screen display and 3x4 optional keypad input. The Samsung S8300 is acclaimed to be one of the better Samsung mobile phones to date. The phone has a vast array of features such as; an 8.0 megapixel camera with video recording(Auto-focus, Face Detection & Smile-Shot) and an Integrated Music Player. @en
http://dbpedia.org/ontology/abstract	"삼성 울트라 터치(Samsung Ultra Touch), 삼성 토코 올트리 에디션(Samsung Tocco Ultra Edition), 삼성 S8300/Samsung S8300은 대한민국에서는 앤디풀 월드와 벨리크로 출시된 삼성전자와의 협력 전화이다. 880mAh의 표준 배터리만 지원하여, 1600mAh 캔버스의 능동형 유기 발광 다이오드(AM OLED) 화면을 적용하여 TFT LCD 대비 반응속도가 1000배 더 빠르고, 전력소모량이 적다. '벨리U'를 내장하였다. 영국의 모바일 쇼어스에서 별 5개의 성적을 기록하였다. @ko
http://dbpedia.org/property/weight	"122.0^^< http://dbpedia.org/datatype/gram >
http://dbpedia.org/property/wikiPageUsesTemplate	http://dbpedia.org/resource/Template:Infobox mobile phone

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

Results Format: [HTML](#) (The CXML output is disabled, see [details](#))

Execution timeout: 0 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](#))

[Run Query](#) [Reset](#)

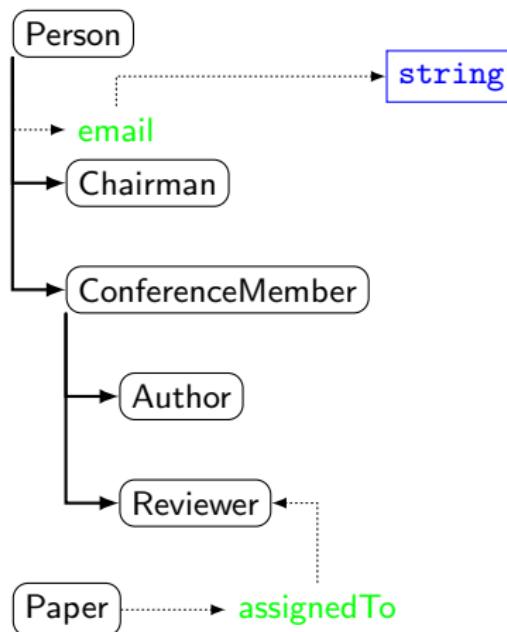
Copyright © 2012 [OpenLink Software](#)
Virtuoso version 06.04.3132 on Linux (x86_64-generic-linux-glibc212-64), Cluster Edition (4 server processes)

Find: < Previous > Next Highlight all Match case

Ontologies : Key ingredients of the semantic web

- ▶ Define the concepts and relationships used to describe and represent an area of concern
- ▶ Define possible constraints on using those concepts and relationships
- ▶ Building blocks for inference techniques on the Semantic Web

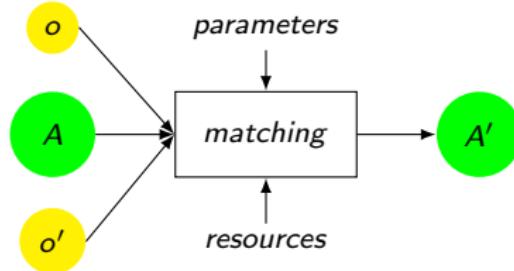
Ontologies



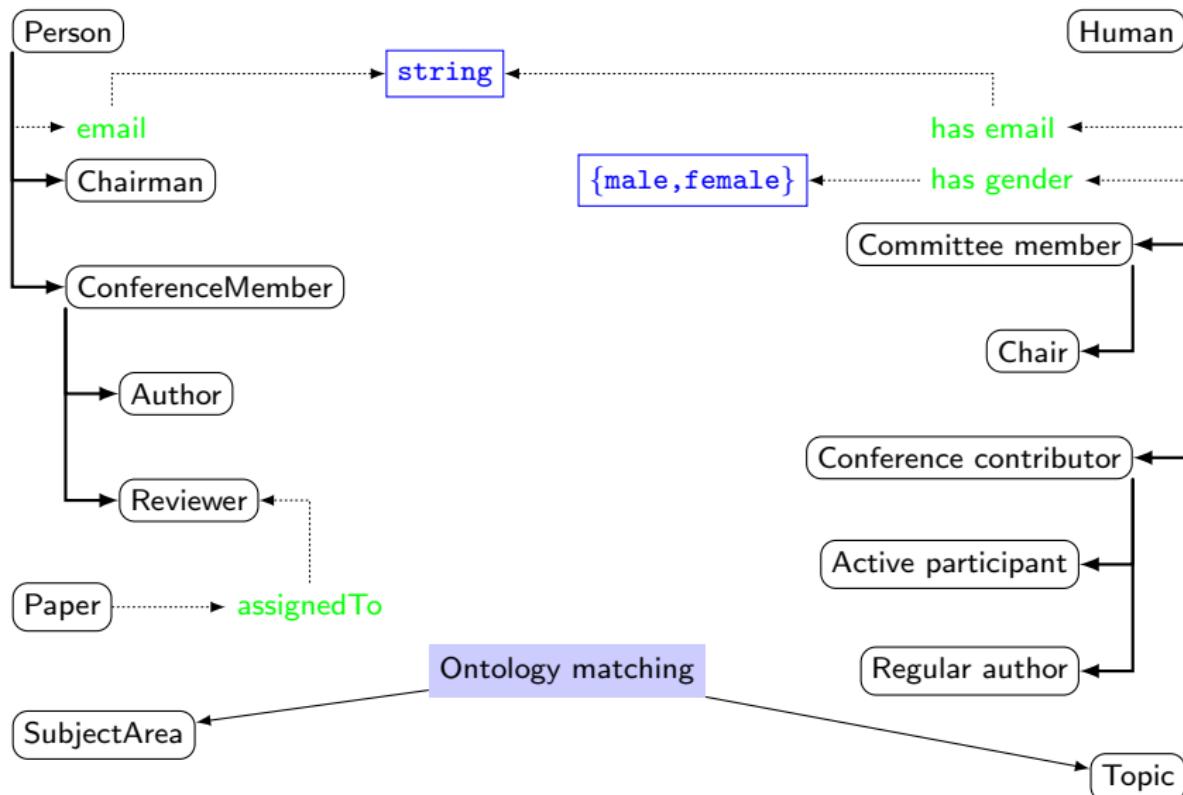
Ontology matching

- ▶ Ontologies shift the heterogeneity problem to a different level
 - ▶ different schemes for web site annotation
 - ▶ different vocabularies for describing data sets
 - ▶ different ontologies describing agents' world
- ▶ For many tasks, ontologies have to be conciliated

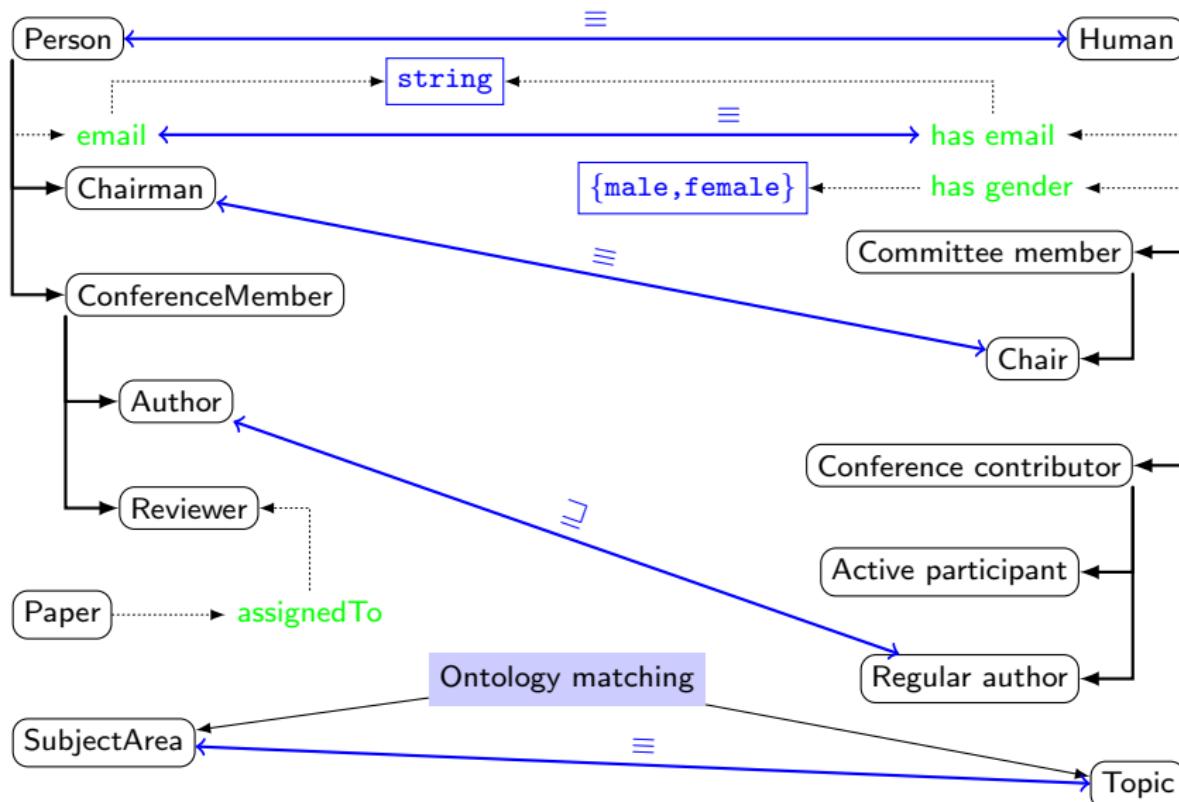
Ontology matching



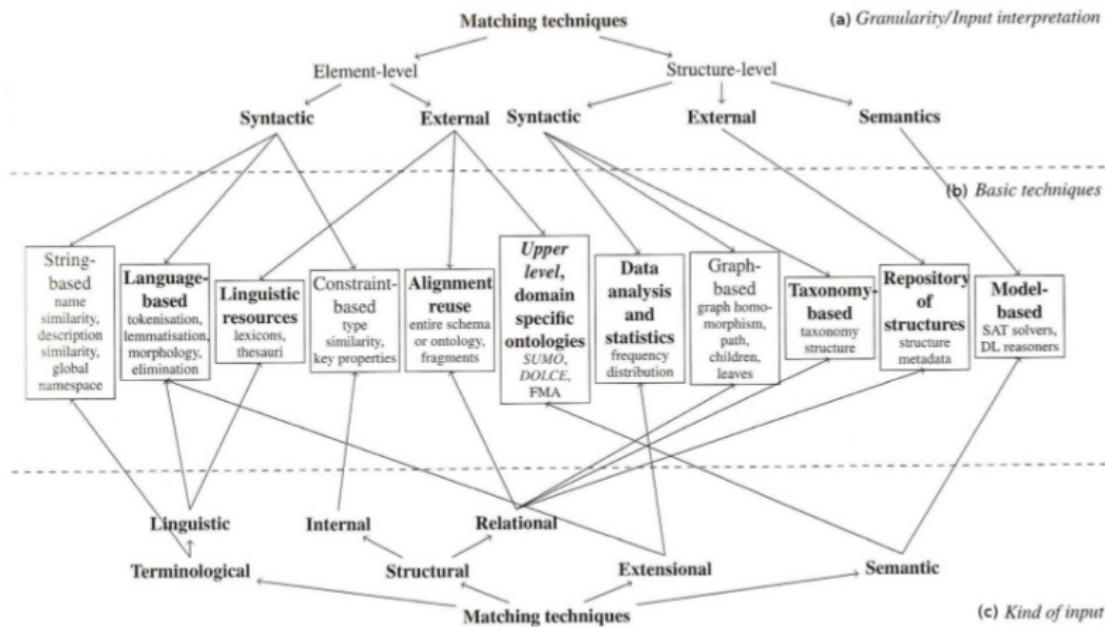
Ontology matching



Ontology matching



Ontology matching techniques



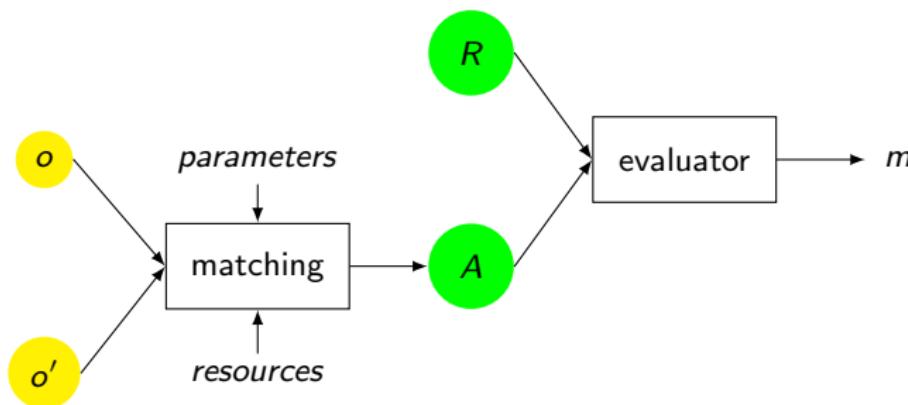
(Euzenat and Shvaiko [2007])

Ontology matching techniques

- ▶ Matching systems rely not on a single approach
- ▶ Diverse ways of exploring the matching problem
- ▶ Suitability of approaches : depends on ontologies/tasks

Ontology matching evaluation

- Objective** Help **developers** of matching systems to improve their systems
Help **users** evaluating the suitability of proposed systems to their needs



Ontology matching evaluation

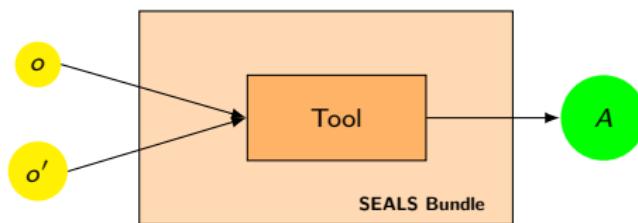
Initiatives Ontology Alignment Evaluation Initiative (OAEI)

Annual campaigns since 2004

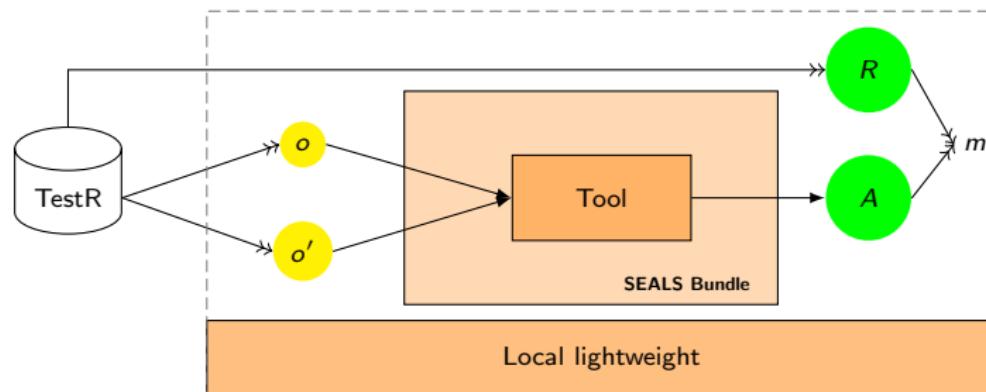
Different tracks from different domains

Automatisation Semantic Evaluation at Large Scale (SEALS)
Benchmark test generator : scalability and difficulty

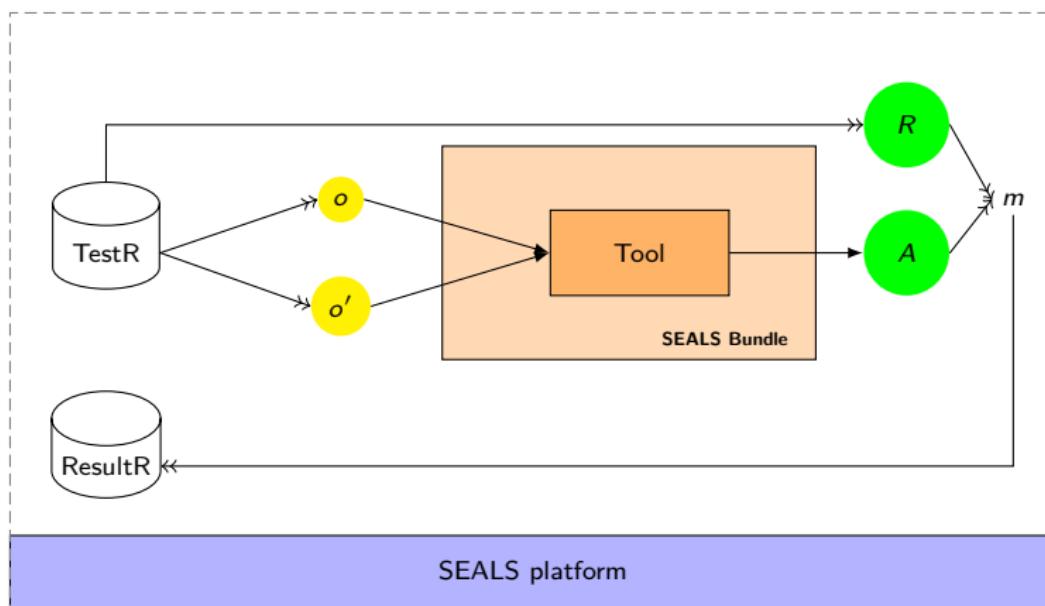
Automatic evaluation in SEALS



Automatic evaluation in SEALS



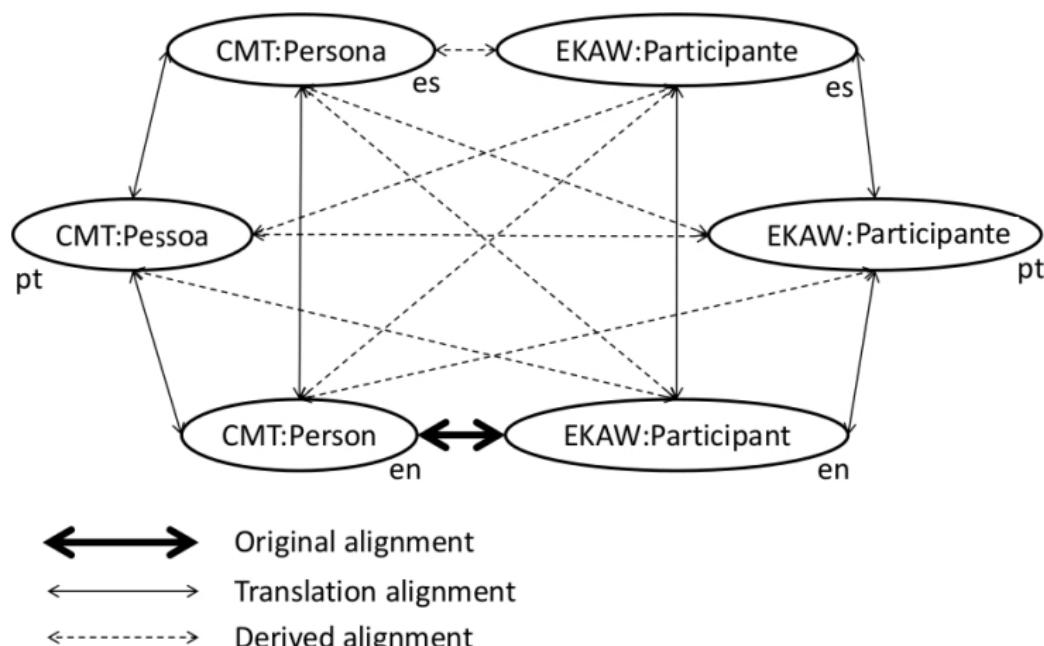
Automatic evaluation in SEALS



Multilingual ontology matching

Benchmark MultiFarm : benchmark for multilingual matching
7 conference ontologies translated in 8 languages :
CZ, CN, DE, EN, ES, DE, FR, RU, PT (+EN)

MultiFarm data set



Background
oooooooo

Ontology matching
ooooooo

Ontology matching evaluation
ooo

Multilingual matching
oo●ooooo

MultiFarm data set

Name	Type	#C	#DP	#OP	DL
Ekaw	Insider	74	0	33	<i>SHIN</i>
Sofsem	Insider	60	18	46	<i>ALCHIF</i>
Sigkdd	Web	49	11	17	<i>ALEI</i>
lasted	Web	140	3	38	<i>ALCIN</i>
ConfTool	Tool	38	23	13	<i>SIN</i>
Cmt	Tool	36	10	49	<i>ALCIN</i>
Edas	Tool	104	20	30	<i>ALCOIN</i>

OAEL 2012 : MultiFarm test set

Subset 5 ontologies (36×25 matching tasks)

Participation 7 participants (out of 22) implementing specific multilingual methods (3 in OAEI 2011.5)

Specific systems ASE, AUTOMSv2, GOMMA, MEDLEY, WeSeE, Wmatch, and YAM++

Strategies Translation, bilingual dictionaries, Wikipedia for extracting inter-language links

OAEI 2012 : MultiFarm test set

Preliminary results of MultiFarm here

MultiFarm

The goal of the MultiFarm track is to evaluate the ability of matcher systems to deal with multilingual ontologies. It serves the purpose of evaluating the strength and weakness of matchers (by being progressive and wide coverage) and measuring the progress of matchers (by being stable and reusable over the years), with a special focus on multilingualism.

- Schedule
- Data sets
- Testing your tool
- Contacts

Schedule

The schedule is that of <http://oaei.ontologymatching.org/2012/>.

Evaluation campaign results

The results of the OAEI 2012, for the MultiFarm track are available [here](#).

Data sets

The MultiFarm data set is composed of a set of ontologies translated in different languages and the corresponding alignments between these ontologies. It is based on the [OntoFarm dataset](#), which has been used successfully for several years in the [Conference track](#) in the OAEI campaigns. By translating the ontologies of the OntoFarm dataset into eight different languages:

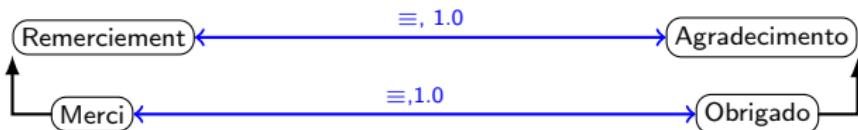
- Chinese (cn)
- Czech (cz)
- Dutch (nl)
- French (fr)
- German (de)

MultiFarm observations

- ▶ Large variation in runtime : from $\cong 15$ minutes to $\cong 17$ hours ;
- ▶ Specific methods work much better than non specific systems (expected behaviour) ;
- ▶ With the exception of Wmatch, all systems use English as pivot language ;
- ▶ Absolute results are still not very good if compared with the results on the original conference test set.

References

- ▶ J. Euzenat and P. Shvaiko. *Ontology Matching*. Springer-Verlag, Heidelberg (DE), 2007.



Questions ?