

# **NERD meets NIF: Lifting NLP Extraction Results to the Linked Data Cloud**

*Giuseppe Rizzo* and Raphaël Troncy

EURECOM, France

Sebastian Hellmann and Martin Bruemmer

Universität Leipzig, Germany

# What is a Named Entity recognition task?

A task that aims to locate and classify the name of a person or an organization, a location, a brand, a product, a numeric expression including time, date, money and percent in a textual document



# NER tools

---

- Standalone software
  - GATE
  - Stanford CoreNLP
  - Temis
  
- Web APIs



# Factual comparison of 10 Web NER tools

	Alchemy API	DBpedia Spotlight	Evri	Extractiv	Lupedia	Open Calais	Saplo	Wikimeta	Yahoo!	Zemanta
Language	EN,FR,GR,IT,PT,RU,SP,SW	EN GR* PT* SP*	EN,IT	EN	EN,FR,IT	EN,FR SP	EN,SW	EN,FR SP	EN	EN
Granularity	OEN	OEN	OED	OEN	OEN	OEN	OED	OEN	OEN	OED
Entity position	N/A	char offset	N/A	word offset	range of chars	char offset	N/A	POS offset	range of chars	N/A
Classification schema	Alchemy	DBpedia FreeBase Scema.org	Evri	DBpedia	DBpedia LinkedMDB	Open Calais	N/A	ESTER	Yahoo	FreeBase
Number of classes	324	320	5	34	319	95	5	7	13	81
Response Format	JSON MicroF XML RDF	HTML JSON RDF XML	HTM L JSO N RDF	HTML JSON RDF XML	HTML JSON RDFa XML	JSON MicroF ormat	JSON	JSON XML	JSON XML	XML JSON RDF
Quota (calls/day)	30000	unl	3000	3000	unl	50000	1333	unl	5000	10000

## What is NERD?

ontology<sup>1</sup>      REST API<sup>2</sup>  
UI<sup>3</sup>



The NERD ontology has been integrated in the NIF project, a EU FP7 in the context of the LOD2: Creating Knowledge out of Interlinked Data

<sup>1</sup> <http://nerd.eurecom.fr/ontology>

<sup>2</sup> <http://nerd.eurecom.fr/api/application.wadl>

<sup>3</sup> <http://nerd.eurecom.fr>

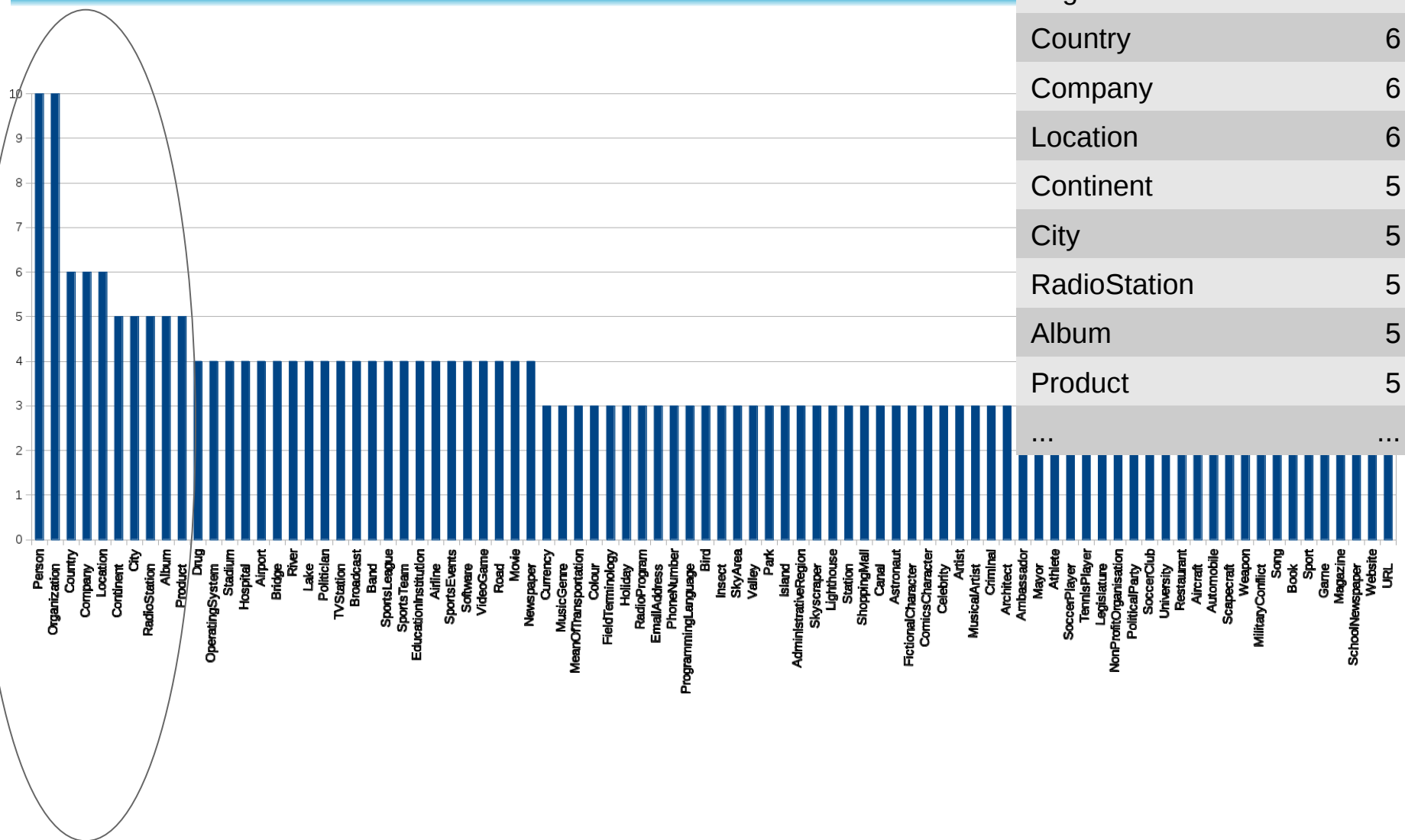
# NERD Ontology

---



Aligned the taxonomies used by  
the extractors

# Building the NERD Ontology



# Ontology alignment validation

5 TED  
talks

	AlchemyAPI	DBpedia Spotlight	Evri	Extractiv	OpenCalais	Zemanta
Person	42	-	10	6	27	4
Organization	15	-	-	-	20	1
Country	16	-	11	1	16	3
City	14	-	3	3	7	-
Time	-	-	-	1	-	-
Number	-	-	-	5	-	-

1000  
NYT  
news  
articles

	AlchemyAPI	DBpedia Spotlight	Evri	Extractiv	OpenCalais	Zemanta
Person	6,246	14	2,698	5,648	5,615	1,069
Organization	2,479	-	900	81	2,538	180
Country	1,727	2	1,382	2,676	1,707	720
City	2,133	-	845	2,046	1,863	-
Time	-	-	-	123	1	-
Number	-	-	-	3,940	-	-

217  
WWW2011  
abstracts

	AlchemyAPI	DBpedia Spotlight	Evri	Extractiv	OpenCalais	Zemanta
Person	17	-	12	6	6	1
Organization	20	-	24	-	5	-
Country	9	-	8	14	7	6
City	4	-	3	8	9	-
Time	-	-	-	-	-	-
Number	-	-	-	184	-	-



# Integration

- Different outputs for the NLP tools (Standalone and Web APIs)

## OpenCalais

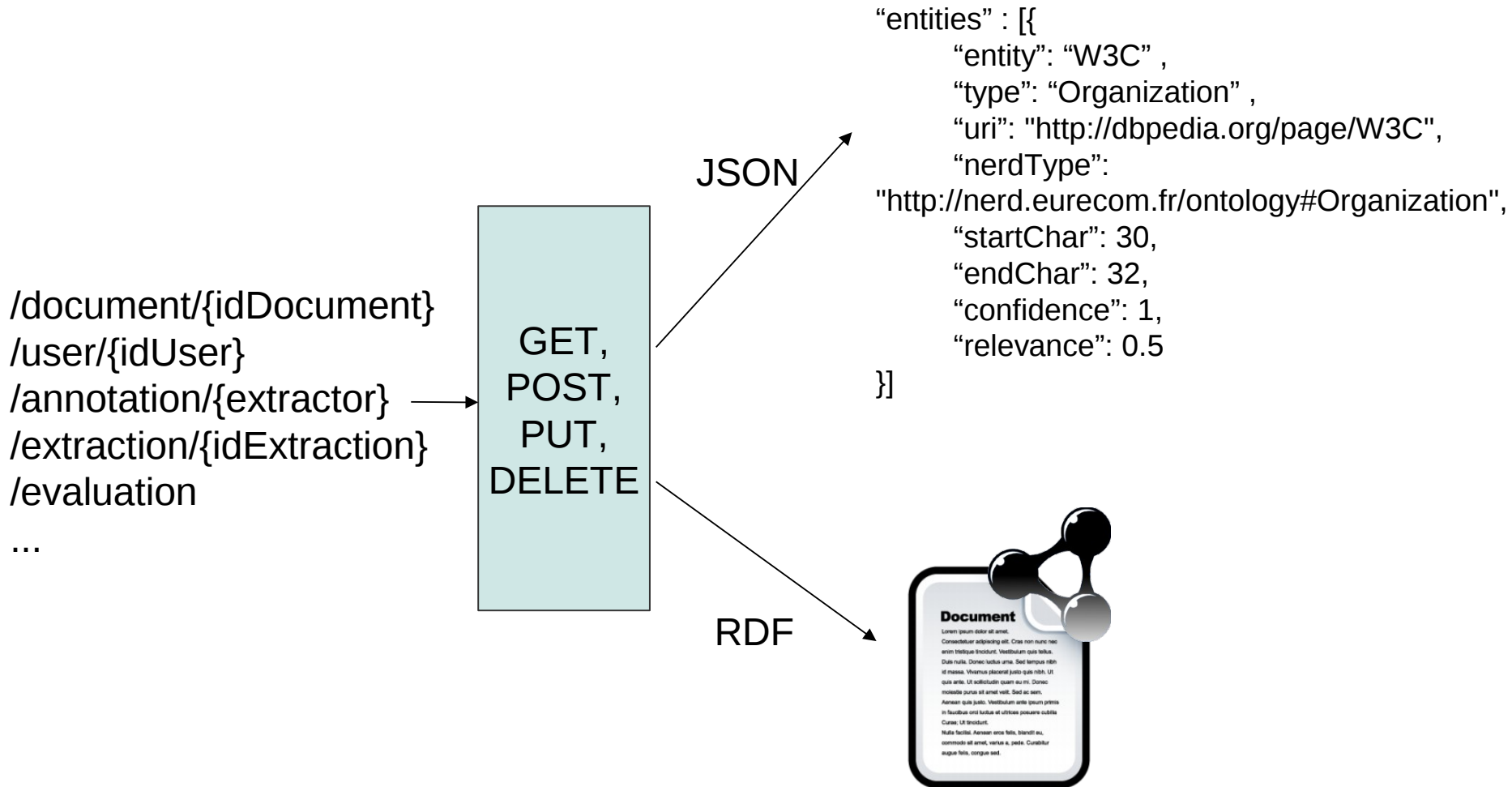
```
"_type": "Organization",  
"name": "North Atlantic Treaty Organization",  
"organizationtype": "governmental civilian",  
"nationality": "N/A",  
"_typeReference":  
  http://s.opencalais.com/1/type/em/e/Organization",  
...
```

## DBpedia Spotlight

```
"@URI": "http://dbpedia.org/resource/DBpedia",  
"@types": "DBpedia:Software,DBpedia:Work"  
"@surfaceForm": "dbpedia",  
"@offset": "0",  
"@support": "11",  
"@similarityScore": "0.2387271374464035",  
...
```

- For integration or reuse manual effort is needed
  - time consuming
  - difficult to track definitions
- NERD creates a sharable JSON/RDF annotation output

# NERD REST API



# Textual annotation

---

Let's consider the URI:

<http://www.w3.org/DesignIssues/LinkedData.html>

The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.....

All the above plus, Use open standards from **W3C** (RDF and SPARQL) to identify things, so that people can point at your stuff

...

entities: {

...

[entity: W3C, startChar: 23107, endChar: 23110],

...

}

# NERD meets NIF



Model documents through a set of strings deferencable within the Web

```
: offset_23107_23110 a str:String ;  
  str:referenceContext :offset_0_26546 .
```

N · E · R · 

Map string to entity

```
: offset_23107_23110 sso:oen dbpedia:W3C .
```



Classification

```
dbpedia:W3C rdf:type nerd:Organization .
```

# NERD User Interface



You are logged in.

[Logout!](#)

[home](#)

[documentation](#)

[ontology](#)

[publication](#)

URI article

Analyze web resource

Raw text

Analyze plain text




© 2012. Eurecom. Sophia-Antipolis

# Conclusions and perspectives

---

NERD UI and REST API 

unified interface for extracting NEs from various type of texts

NERD ontology 

common schema for entity classification

NERD & NIF 

lift the extraction annotation results to the LOD cloud

Systematic comparison for the NE extraction and classification tasks:

ETAPE corpus

CoNLL 2003 corpus

Combining several extractions to improve the strengths of a single tool

# Thanks for your time and your attention

---

N·E·R·D

<http://nerd.eurecom.fr>



@giusepperizzo @rtroncy #nerd



slideshare

<http://www.slideshare.net/giusepperizzo>