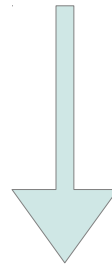


# **NERD@ETAPE2012: results of the EN task**

# N·E·R·1



- ✓ Web-service based extraction
- ✓ Extraction results = set of (named entity, type, URI)

# Different languages support

	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	Extractiv	DBpedia LinkedM DB	Open Calais	Saplo	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	HTML JSON RDF	HTML JSON RDF XML	HTML JSON RDFa XML	JSON MicroFormat	JSON	JSON XML	JSON XML	XML JSON RDF
Quota (calls/day)	30000	unl	3000	3000	unl	50000	1333	unl	5000	10000

# Different classification schema

	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	Extractiv	DBpedia LinkedM DB	Open Calais	Saplo	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	HTML JSON 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

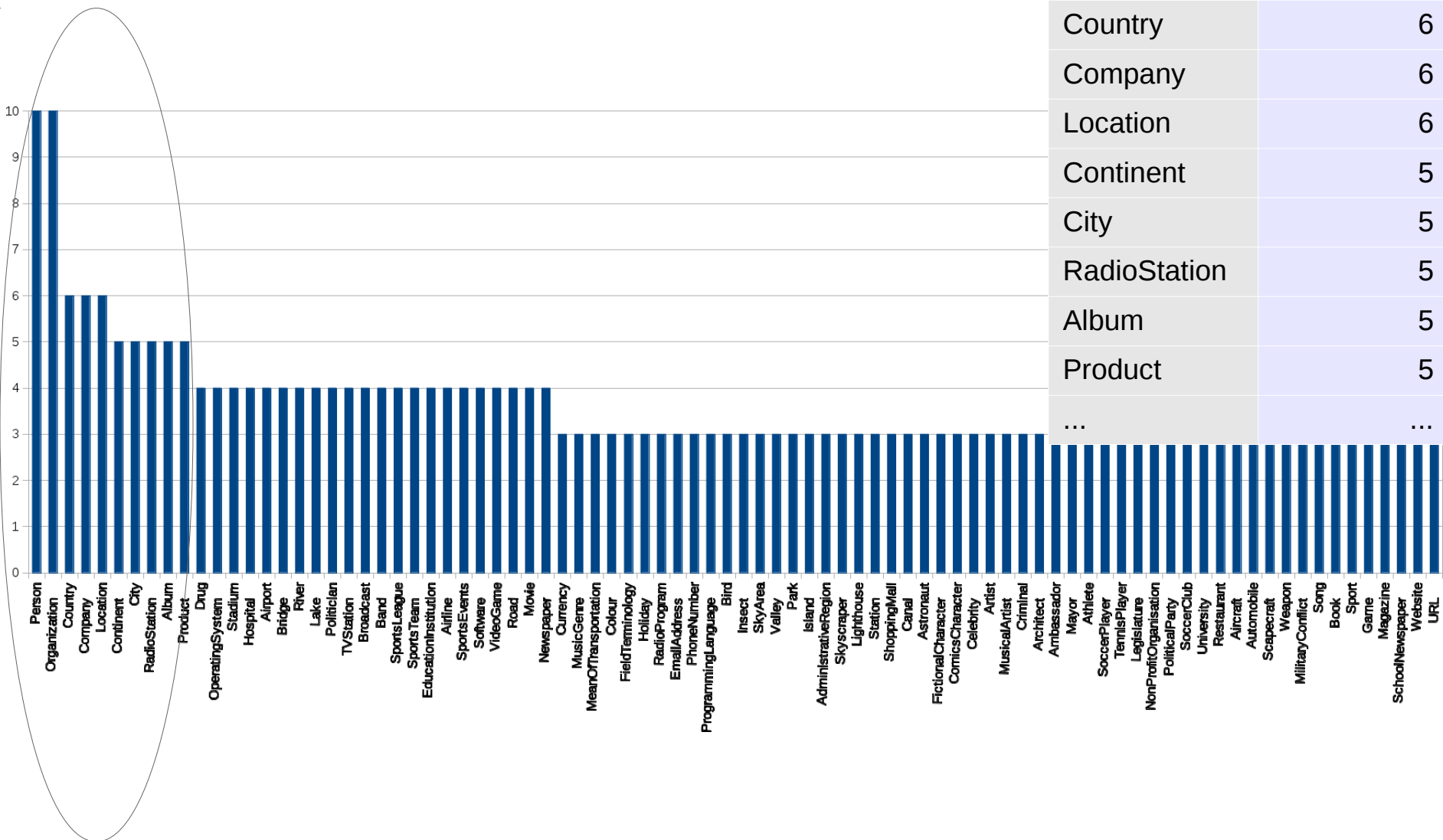
# NERD Ontology

---



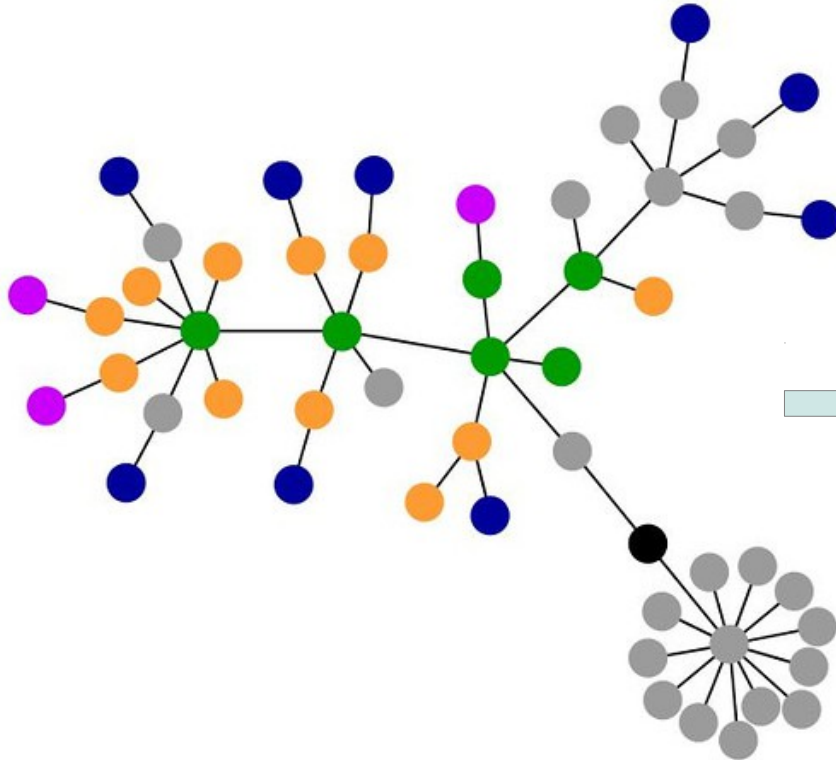
Aligned the taxonomies used by  
the extractors

# Building the NERD Ontology

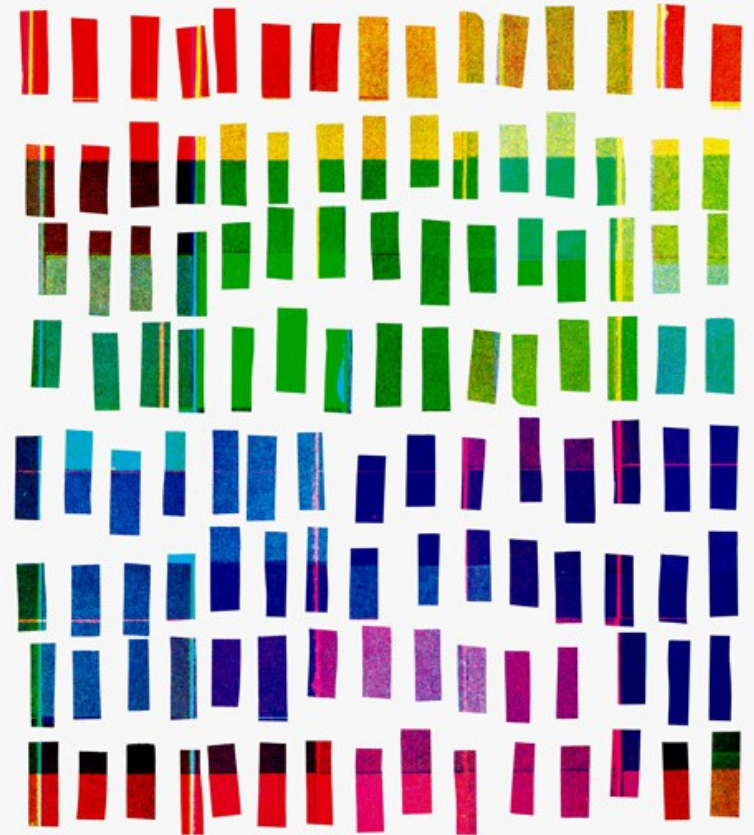
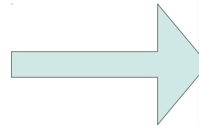


NERD type	Occurrence
Person	10
Organization	10
Country	6
Company	6
Location	6
Continent	5
City	5
RadioStation	5
Album	5
Product	5
...	...

# Type mapping

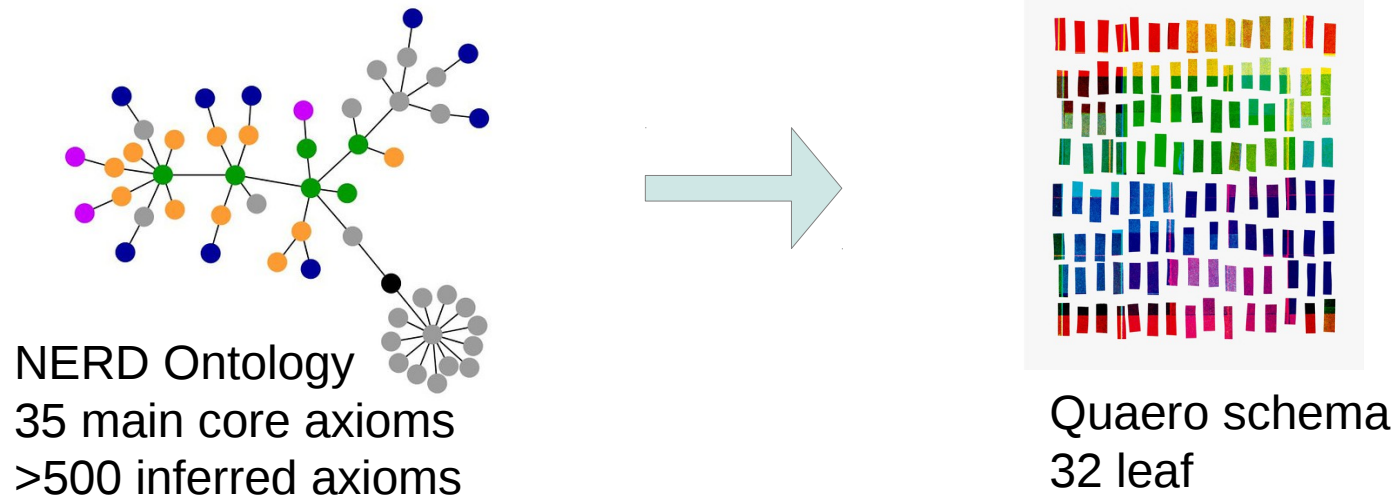


NERD Ontology  
35 main core axioms  
>500 inferred axioms



Quaero schema  
32 leaf

# Alignment mismatch



426 final NERD axioms mapped to 32 leaf,  
loosing the entire semantics about the root

Q. is pers.ind an inferred axiom of Person?  
A. no

Mismatch between the NERD ontology and the  
Quaero schema



# NERD strategy at ETAPE

---

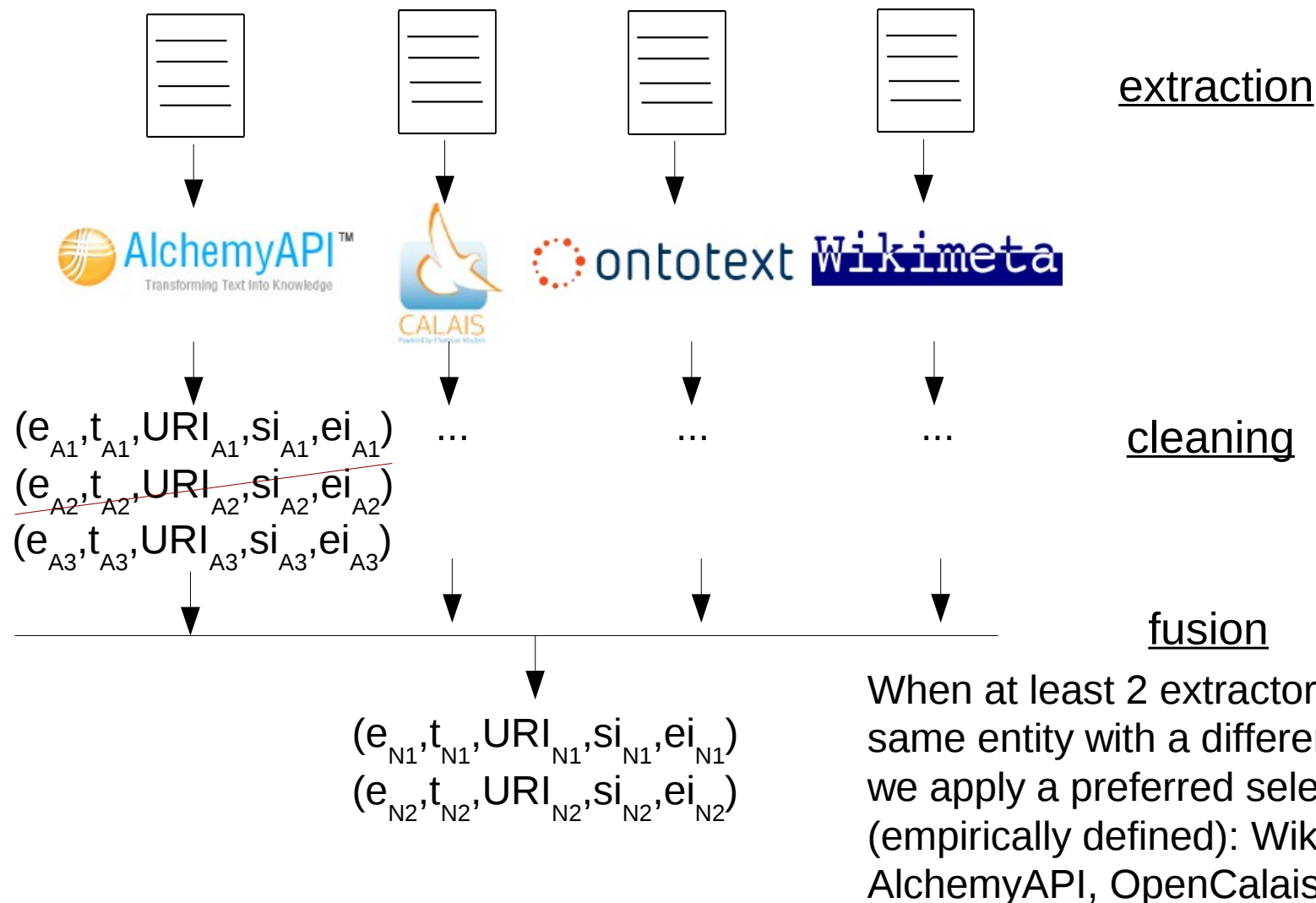
## Baselines:

1. Wikimeta
2. AlchemyAPI
3. Lupedia
4. OpenCalais

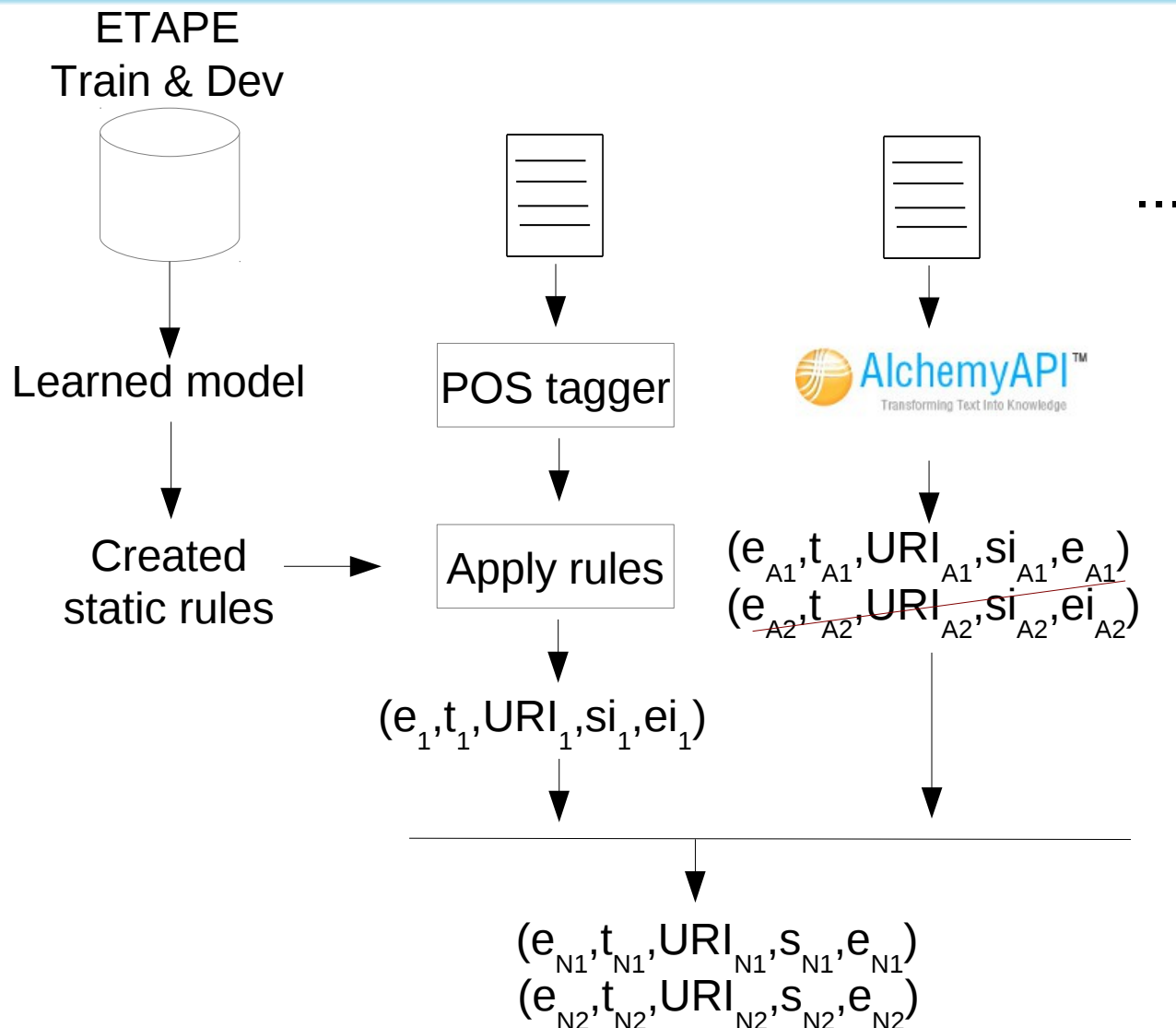
## 2 runs:

1. combined
2. combined+

# combined



# combined+



## fusion

Conflicts handled by  
priority selection: own,  
Wikimeta, AlchemyAPI,  
OpenCalais, Lupedia

# NERD Global results

---

	SLR	prec	recall	F-measure	%correct
combined	<b>86.85%</b>	<b>35.31%</b>	17.69%	<b>23.44%</b>	17.69%
combined+	188.81%	15.13%	<b>28.40%</b>	19.45%	<b>28.40%</b>

combined+ : Eval corpus differs substantially from the Train & Dev corpora. The static rules do not fit well the Eval corpora and they introduce noise.

# Per-extractor results

	SLR	prec	recall	F-measure	%correct
alchemyapi	37.71%	<b>47.95%</b>	5.45%	9.68%	5.45%
lupedia	39.49%	22.87%	1.56%	2.91%	1.56%
opencalais	<b>37.47%</b>	41.69%	3.53%	6.49%	3.53%
wikimeta	36.67%	19.40%	4.25%	6.95%	4.25%
combined (nerd)	86.85%	35.31%	17.69%	<b>23.44%</b>	17.69%
combined+ (nerd+)	188.81%	15.13%	<b>28.40%</b>	19.45%	<b>28.40%</b>

# The tree of “Person” leaf

	pers.ind			pers.coll		
	prec	recall	F-measure	prec	recall	F-measure
combined	<b>78.11%</b>	57.46%	<b>65.51%</b>	0.00%	0.00%	0.00%
combined+	48.55%	<b>70.40%</b>	56.25%	<b>15.95%</b>	<b>44.11%</b>	<b>22.21%</b>

OUTLIER = 1 pers.other type defined in the GT @ V8\_LaPlaceDuVillage\_2011-05-12\_17280. This leaf doesn't exist in the Quaero schema (ref. September 2011)

# per-type results (I)

	person			organization			location		
	prec	recall	F-measure	prec	recall	F-measure	prec	recall	F-measure
combined	<b>39.06%</b>	28.73%	32.76%	8.55%	13.67%	9.40%	26.84%	27.22%	23.99%
combined+	32.25%	<b>57.25%</b>	<b>39.23%</b>	<b>18.45%</b>	<b>31.13%</b>	<b>21.28%</b>	<b>28.12%</b>	<b>30.01%</b>	<b>24.10%</b>

# The tree of “Organization” leaf

	org.ent			org.adm		
	prec	recall	F-measure	prec	recall	F-measure
combined	16.33%	27.15%	18.50%	0.76%	0.19%	0.30%
combined+	<b>19.31%</b>	<b>40.43%</b>	<b>25.37%</b>	<b>17.59%</b>	<b>21.83%</b>	<b>17.19%</b>



# per-type results (I)

	person			organization			location		
	prec	recall	F-measure	prec	recall	F-measure	prec	recall	F-measure
combined	<b>39.06%</b>	28.73%	32.76%	8.55%	13.67%	9.40%	26.84%	27.22%	23.99%
combined+	32.25%	<b>57.25%</b>	<b>39.23%</b>	<b>18.45%</b>	<b>31.13%</b>	<b>21.28%</b>	<b>28.12%</b>	<b>30.01%</b>	<b>24.10%</b>

# The tree of “Location” leaf

	loc.adm.town			loc.adm.reg			loc.adm.nat		
	prec	recall	F-measure	prec	recall	F-measure	prec	recall	F-measure
combined	50.88%	46.03%	42.54%	<b>13.82%</b>	35.79%	17.13%	55.50%	<b>64.64%</b>	<b>58.34%</b>
combined+	<b>55.11%</b>	<b>49.65%</b>	<b>46.47%</b>	12.38%	<b>45.79%</b>	<b>17.45%</b>	<b>56.99%</b>	50.54%	52.39%

	loc.adm.sup			loc.phys.geo			loc.phys.hydro		
	prec	recall	F-measure	prec	recall	F-measure	prec	recall	F-measure
combined	<b>36.05%</b>	<b>53.03%</b>	<b>41.20%</b>	<b>41.43%</b>	18.84%	24.76%	<b>16.67%</b>	<b>8.33%</b>	<b>11.12%</b>
combined+	0.84%	36.91%	1.64%	37.50%	<b>18.99%</b>	<b>25.03%</b>	12.50%	6.25%	8.34%

OUTLIER = 2 types of loc.oro (they should be loc.phys.oro) @EST2BC\_FRE-FR-20101007\_21\_52\_FINTER\_DEBATE & @LCP\_TopQuestions\_2011-05-18\_000400

# per-type results (I)

	person			organization			location		
	prec	recall	F-measure	prec	recall	F-measure	prec	recall	F-measure
combined	<b>39.06%</b>	28.73%	32.76%	8.55%	13.67%	9.40%	26.84%	27.22%	23.99%
combined+	32.25%	<b>57.25%</b>	<b>39.23%</b>	<b>18.45%</b>	<b>31.13%</b>	<b>21.28%</b>	<b>28.12%</b>	<b>30.01%</b>	<b>24.10%</b>

# The tree of “Time” leaf

	time.date.abs			time.date.rel			time.hour.abs			time.hour.rel		
	prec	recall	F-measure	prec	recall	F-measure	prec	recall	F-measure	prec	recall	F-measure
combined	9.81 %	25.85 %	13.73 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %
combined+	<b>11.11 %</b>	<b>42.16 %</b>	<b>16.69 %</b>	<b>51.85 %</b>	<b>53.17 %</b>	<b>51.63 %</b>	<b>12.99 %</b>	<b>15.61 %</b>	<b>14.05 %</b>	<b>9.94 %</b>	<b>39.10 %</b>	<b>14.65 %</b>

OUTLIER = 2 types of prod.rule @BFMTV\_BFMStory\_2011-05-31\_175900 & @LCP\_EntreLesLignes\_2011-05-06\_192800 & @LCP\_EntreLesLignes\_2011-05-13\_192800; 1 type of prod.unk @LCP\_TopQuestions\_2011-05-18\_000400

# per-type results (II)

	time			prod			func		
	prec	recall	F-measure	prec	recall	F-measure	prec	recall	F-measure
combined	2.45%	6.46%	3.43%	7.35%	2.98%	4.03%	16.55%	5.68%	8.30%
combined+	<b>21.47%</b>	<b>37.51%</b>	<b>24.26%</b>	<b>12.06%</b>	<b>10.66%</b>	<b>9.76%</b>	<b>34.14%</b>	<b>30.70%</b>	<b>27.06%</b>

# The tree of “Product” leaf

	prod.object			prod.art			prod.media			prod.fin		
	prec	recall	F-measure	prec	recall	F-measure	prec	recall	F-measure	prec	recall	F-measure
combined	<b>2.78</b> %	<b>0.64</b> %	<b>1.04</b> %	<b>10.05</b> %	<b>4.11</b> %	<b>5.13</b> %	<b>42.18</b> %	17.32 %	23.47 %	11.11 %	4.77 %	6.67 %
combined+	2.56 %	0.59 %	0.96 %	9.05 %	3.79 %	4.53 %	39.89 %	<b>44.13</b> %	<b>39.39</b> %	<b>34.51</b> %	<b>19.91</b> %	<b>22.32</b> %

# per-type results (II)

	time			prod			func		
	prec	recall	F-measure	prec	recall	F-measure	prec	recall	F-measure
combined	2.45%	6.46%	3.43%	7.35%	2.98%	4.03%	16.55%	5.68%	8.30%
combined+	<b>21.47%</b>	<b>37.51%</b>	<b>24.26%</b>	<b>12.06%</b>	<b>10.66%</b>	<b>9.76%</b>	<b>34.14%</b>	<b>30.70%</b>	<b>27.06%</b>

# The tree of “Function” leaf

	func.ind			func.coll		
	prec	recall	F-measure	prec	recall	F-measure
combined	<b>33.09%</b>	11.36%	16.61%	0.00%	0.00%	0.00%
combined+	29.63%	<b>36.21%</b>	<b>30.02%</b>	<b>38.65%</b>	<b>25.18%</b>	<b>24.11%</b>



# per-type results (II)

	time			prod			func		
	prec	recall	F-measure	prec	recall	F-measure	prec	recall	F-measure
combined	2.45%	6.46%	3.43%	7.35%	2.98%	4.03%	16.55%	5.68%	8.30%
combined+	<b>21.47%</b>	<b>37.51%</b>	<b>24.26%</b>	<b>12.06%</b>	<b>10.66%</b>	<b>9.76%</b>	<b>34.14%</b>	<b>30.70%</b>	<b>27.06%</b>

# Amount

---

	prec	recall	F-measure
combined	<b>49.87%</b>	13.83%	<b>20.98%</b>
combined+	5.68%	<b>22.92%</b>	8.87%

# Lessons learned

---

## ➤ **ontology**

- deeper classification schemes reduce the alignment mismatch from the NERD ontology and the Quaero schema.

## ➤ **combined**

- it outperforms the baselines
- weighted extractor selection according to the strengths of the extractors. e.g. AlchemyAPI pers.ind, while Wikimeta amount

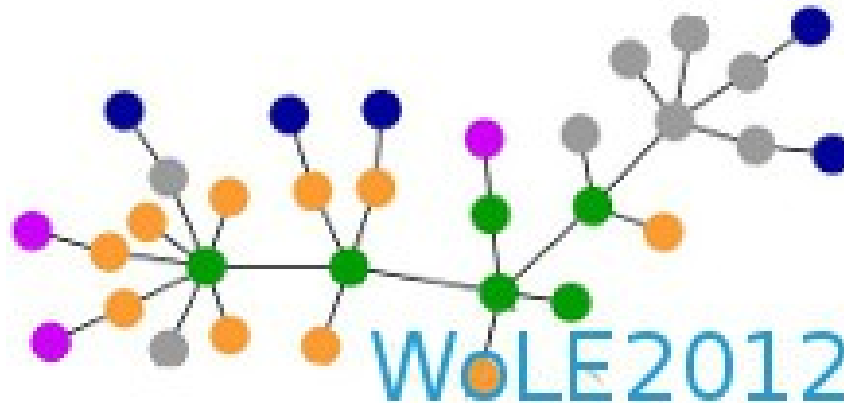
## ➤ **combined+**

- it outperforms the combined (for pers.ind it has a lower precision due to the noise introduced)
- add a probabilistic approach

# Announcement

---

**WoLE2012 Workshop in conjunction with the  
ISWC2012 conference**



**ISWC2012 Workshop**

<http://wole2012.eurecom.fr>