A Comparison of Knowledge Extraction Tools for the Semantic Web

Aldo Gangemi

aldo.gangemi @cnr.it , @lipn.univ-paris13.fr

LIPN, Université Paris 13, UMR CNRS, Sorbonne Cité

Semantic Technology Lab, ISTC-CNR, Rome, Italy

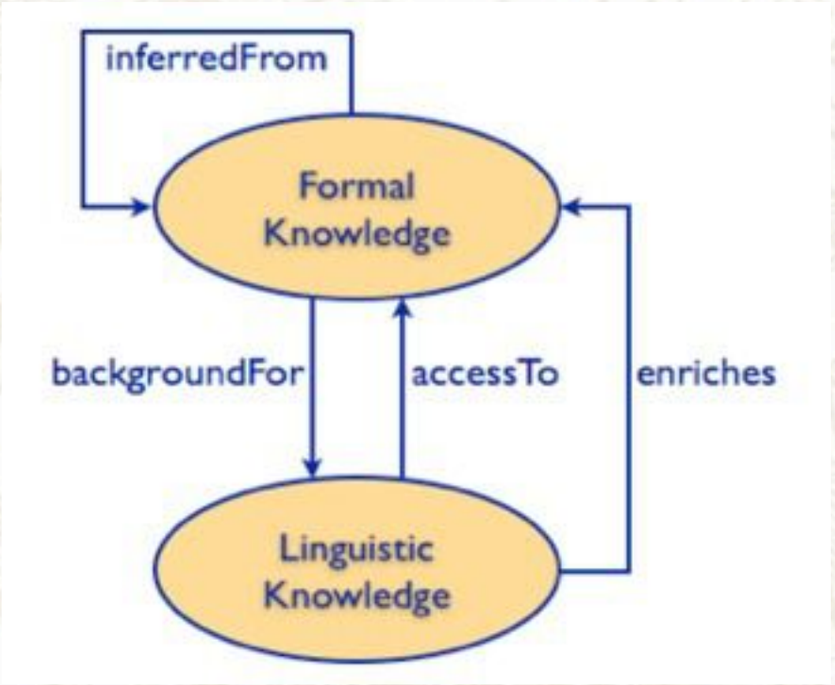
Objective and results

- A landscape analysis of the current tools for Knowledge Extraction from text (KE), when applied on the Semantic Web (SW)
- An assessment of KE2SW capabilities of 16 tools on 14 basic tasks

Knowledge Extraction and the SW

- Historically SW mainly worked on *ontology learning*
  - unconvincing results: sparseness, core knowledge difficult to catch, etc. (cf. Coppola et al. 2009, Blomqvist, 2009)
  - natural language understanding known to be an AI-complete problem
- The paradigm of *machine reading* (Etzioni, 2006) fits the lightweight and/or data-driven trend of current SW
- More important, *semantic technologies* in the large need hybridization

QA, NL querying on LD, full text search jointly with queries, ...



Google Knowledge Graph

IBM Watson

OIE, NELL, BabelNet, ...

Trend against the “semiotic divide”

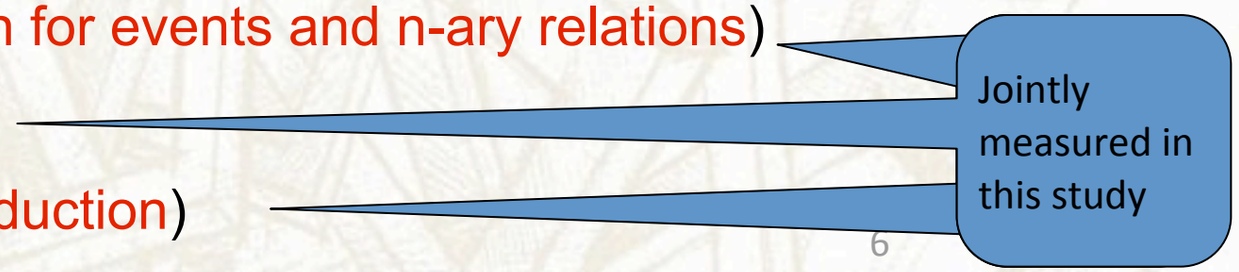
- Linguistic semantics is based typically on informal intensional relations
- Logical semantics is typically based on formal extensional axioms
- KE tools are entering SW production with LD population and ontology evolution frameworks
  - Specially in the form of mashups or merging platforms (FOX, NER, Stanbol), and interoperability protocols (NIF, FISE)
- Still a large gap and heterogeneous solutions
  - Cf. W3C Community Group Ontology-Lexicon
  - 2 main use cases: linguistic linked data vs. ontology design/evolution via machine reading

Basic vs. application tasks

- Traditionally, NLP tasks distinguished into:
  - basic (e.g. named entity recognition)
  - applied (e.g. question answering)
- When moving to the SW:
  - basic (e.g. class induction)
  - applied (e.g. NL querying of linked data)

Basic tasks in NLP and SW

- topic extraction (recognition of specific topics, e.g. *individuals from the range of the property dc:subject*)
- named entity recognition (*individual induction*)
- named entity resolution (*identity resolution for individuals*)
- terminology extraction (*induction of constants pertinent to a domain, typically for classes or properties*)
  - (a) class induction
  - (b) property induction
- sense tagging (≈ class membership induction)
- sense disambiguation (≈ identity resolution for classes)
- taxonomy induction (≈ subclass relation induction)
- (non-taxonomic, non-role, binary) relation extraction (property assertion - fact - induction)
- event detection (≈ n-ary relationship - token - induction)
- semantic role labeling (≈ property induction for events and n-ary relations)
- coreference (coreference of individuals)
- frame detection (≈ n-ary relation - type - induction)



Sample translation table

Topic	**<Text> dc:subject <Topic>**
Named entity	**owl:NamedIndividual**
Entity resolution (NE)	**<namedIndividual$_i$> owl:sameAs <namedIndividual$_j$>**
Entity coreference	**<namedIndividual$_i$> owl:sameAs <namedIndividual$_j$>**
Term	**owl:Class || owl:ObjectProperty || owl:DatatypeProperty**
Sense tag	**owl:NamedIndividual rdf:type owl:Class**
Sense disambiguation (classes)	**owl:Class$_i$ owl:equivalentClass owl:Class$_j$**
Taxonomy (subclasses)	**owl:Class$_i$ owl:subClassOf owl:Class$_j$**
Extracted (binary) relationship	**<namedIndividual$_i$> owl:ObjectProperty || owl:DatatypeProperty <namedIndividual$_j$>**
Semantic role	**owl:ObjectProperty || owl:DatatypeProperty**
Event	**<Event> rdf:type <Event.type>**
Frame	**<Event.type> owl:subClassOf <Frame> . <Event> <semrole$_i$> <Entity$_j$>**
Quantified expression	**owl:Restriction**
Linked entities	**owl:sameAs || owl:differentFrom || ...**
Conjunct of individuals	**owl:NamedIndividual <memberOf> <complexObject>**
Disjunction of individuals	**owl:NamedIndividual <memberOf> <collection>**
Factual entailment	**<Event$_1$> <dependency> <Event$_2$>**

The assumptions

- A landscape analysis, no pretense of completeness
- Mapping between intensional and extensional semantics
  - At the level of both tasks and output
- Distinction between pure processing of text, and alignment of processing results to background knowledge
  - E.g. *recognition* vs. *resolution*
- Information measures of the results on basic tasks
  - Precision, Recall, F-measure, Accuracy
- Relative ground truth, see below

What tools?

- The tools selected share certain characteristics that make them a low hanging fruit for our analysis. They are:
  - available as easily installable or downloadable code, web applications, or APIs
  - at least in public demo form
  - oriented to open domain information extraction: not dependent on training to a specific domain
- Licensing is not addressed for this study
- They can either produce SW output, or not
  - we have applied default assumptions on how to convert the output

ReVerb

Open Information Extraction Software

N·E·R·D

NAMED ENTITY RECOGNITION AND DISAMBIGUATION

Part of the
KnowItAll Project

DBpedia Spotlight

Wikimeta

Homepage
HOME

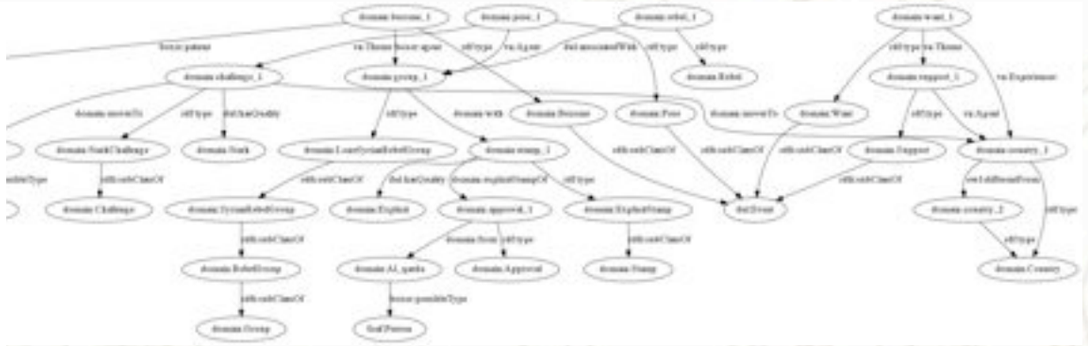
V 1.7 – 31 Jan 2012
Full labeling in 1 s (275 words), language is EN

The lone Syrian [org] rebel group with an explicit stamp of approval from Al Qae
forces, posing a stark challenge to the United States [loc.admi] and other countr
Money flows to the group, the Nusra Front [org] , from like-minded donors abro
skill to storm fortified positions and lead other battalions to capture military bas
weapons and attract more fighters.

The group is a direct offshoot of Al Qaeda in Iraq [org.com], Iraqi [org] officials a
fighters and weapons. " This is just a simple way of returning the favor to our Sy

STLab - tools

Fred



TAG·ME

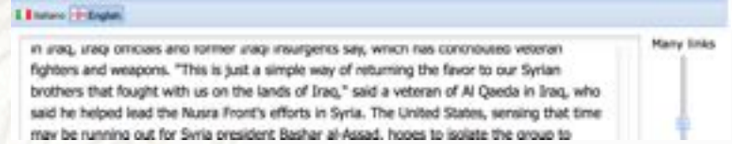
TAGME is a powerful tool that is able to identify on-the-fly meaningful short-phrases (called "spots") in an unstructured text and link them to a pertinent Wikipedia page in a fast and effective way. This annotation process has implications which go far beyond the enrichment of the text with explanatory links because it concerns with the contextualization and, in some way, the understanding of the text.
Try **TAGME** now!

You can play with the demo interface below or check the TAGME RESTful API we are currently supporting.

Currently **TAGME** is available in English and in Italian and it is based on Wikipedia snapshots of July, 2012.

NEWS! As of August 2012, new RESTful functions are available and new advanced parameters can be used. For instance, you can compute semantic relatedness between topics identified by TAGME, or enable the special parser for Twitter messages. Check the RESTful API page for further details.

Developed by Paolo Ferragina and Ugo Scaiella at A³ Lab
Dipartimento di Informatica, University of Pisa



CiceroServer: English

LANGUAGE COMPUTER

ENTITIES

- PERSON (4)
- LOCATION (15)
- ORGANIZATION (6)
- OTHER (6)
- NUMERIC (1)

RELATIONS

- NATIONALITY (1)
- OCCUPATION (1)
- QUOTATION (1)
- GENERIC (15)

DOCUMENT FILE:
(details)

The lone Syrian rebel group with an explicit stamp of approval from Al Qaeda has become one of the uprising most effective fighting forces, posing a stark challenge to the United States and other countries that want to support the rebels but not Islamic extremists. Money flows to the group, the Nusra Front, from like-minded donors abroad. Its fighters, a small minority of the rebels, have the boldness and skill to storm fortified positions and lead other battalions to capture military bases and oil fields. As their successes mount, they gather more weapons and attract more fighters. The group is a direct offshoot of Al Qaeda in Iraq, Iraqi officials and former Iraqi insurgents say, which has contributed veteran fighters and weapons. "This is just a simple way of returning the favor to our Syrian brothers that fought with us on the lands of Iraq," said a veteran of Al Qaeda in Iraq, who said he helped lead the Nusra Front's efforts in Syria. The United States, sensing that time may be running out for Syria president Bashar al-Assad, hopes to isolate the group to prevent it from inheriting Syria, or fighting on after Mr. Assad's fall to pursue its goal of an Islamic state. As the United States pushes the Syrian opposition to organize a viable alternative government, it plans to blacklist the Nusra Front as a terrorist organization, making it illegal for Americans to have financial dealings with the group and prompting similar sanctions from Europe.

ENTITY TYPE KEY

- PERSON
- LOCATION
- ORGANIZATION
- CONTACT_INFO
- OTHER
- DATE/TIME
- NUMERIC

Mouse-over entities and relations to see more detailed type information about them.

Additional information about these types can be found in our online documentation (entities, relations)

DETAILS

Disambiguation Method:

Parameters: (defaults should be OK)

Entities Type Filters:

Mention Extraction:

The lone Syrian rebel group with an explicit stamp of approval from Al Qaeda has become one of the uprising most effective fighting forces, posing a stark challenge to the United States and other countries that

Examples

Input Type:TEXT Overall runtime:1m, 5s, 454ms

The lone [Syria|Syrian] rebel group with an explicit stamp of approval from [Al-Qaeda|Al Qaeda] has become one of the uprising most effective fighting forces, posing a stark challenge to the [United States|United States] and other countries that want to support the rebels but not [Islam|Islamic] extremists. Money flows to the group, the [--NME--|Nusra Front], from like-minded donors abroad. Its fighters, a small minority of the rebels, have the boldness and skill to storm fortified positions and lead other battalions to capture military bases and oil fields. As their successes mount, they gather more weapons and attract more fighters. The group is a direct offshoot of [Al-Qaeda|Al Qaeda] in [Iraq|Iraq], [Iraq|Iraqi] officials and former [Iraq|Iraqi] insurgents say, which has contributed veteran fighters and weapons. "This is just a simple way of returning the favor to our [Syria|Syrian] brothers that fought with us on the lands of [Iraq|Iraq]," said a veteran of [Al-Qaeda|Al Qaeda] in [Iraq|Iraq], who said he helped lead the [--NME--|Nusra Front]'s efforts in [Syria|Syria]. The [United States|United States], sensing that time may be running out for [Syria|Syria] president [Bashar al-Assad|Bashar al-Assad], hopes to isolate the group to prevent it from inheriting [Syria|Syria], or fighting on after Mr. [Bashar al-Assad|Assad]'s fall to pursue its goal of an [Islam|Islamic] state. As the [United States|United States] pushes the [Syria|Syrian] opposition to organize a viable alternative government, it plans to blacklist the [--NME--|Nusra Front] as a terrorist organization, making it illegal for [United States|Americans] to have financial dealings with the group and prompting similar sanctions from [Europe|Europe].

YAGOTypes

CALAIS
Powered by Thomson Reuters

Show RDF | Entry Page

Topics:

Politics	100%
War Conflict	100%

Social Tags:

War Conflict	★★★
Politics	★★★

The lone Syrian reb
countries that want
oldness and skill to
p is a direct offshoo
yrian brothers that
e running out for S
United States pu
dealings with the g

AlchemyAPI™

apache stanbol™

Disclaimer

- Certain tools can be configured (e.g. in terms of confidence or algorithm to be used)
  - We have sticked to default configurations, even if this choice has penalized some of them (in particular Apache Stanbol and partly DBpedia Spotlight)
- Certain tools evolve and may give different result now (FRED, Alchemy, ...)
- More tools are entering the landscape (Tagme, Ollie, TerMine, ...)

Summary of capabilities

Tool/Task	Topics	NER	NE-RS	TE	TE-RS	Senses	Taxo	Rel	Roles	Events	Frames +SRL
AIDA	-	+	+	-	-	+	-	-	-	-	-
Alchemy	+	+	-	+	-	+	-	+	-	-	-
Apache Stanbol	-	+	+	-	-	+	-	-	-	-	-
CiceroLite	-	+	+	+	+	+	-	+	+	+	+
DB Spotlight	-	+	+	-	-	+	-	-	-	-	-
FOX	+	+	+	+	+	+	-	-	-	-	-
FRED	-	+	+	+	+	+	+	+	+	+	+
NERD	-	+	+	-	-	+	-	-	-	-	-
Ollie	-	-	-	-	-	-	-	+	-	-	-
Open Calais	+	+	-	-	-	+	-	-	-	+	-
PoolParty KD	+	-	-	-	-	-	-	-	-	-	-
ReVerb	-	-	-	-	-	-	-	+	-	-	-
Semiosearch	-	-	+	-	+	-	-	-	-	-	-
Tagme	-	+	+	+	+	-	-	-	-	-	-
Wikimeta	-	+	-	+	+	+	-	-	-	-	-
Zemanta	-	+	-	-	-	+	-	-	-	-	-

The test text

A 1500-character text: the upper limit for several tested tools in their web application form

The lone Syrian rebel group with an explicit stamp of approval from Al Qaeda has become one of the uprising most effective fighting forces, posing a stark challenge to the United States and other countries that want to support the rebels but not Islamic extremists. Money flows to the group, the Nusra Front, from like-minded donors abroad. Its fighters, a small minority of the rebels, have the boldness and skill to storm fortified positions and lead other battalions to capture military bases and oil fields. As their successes mount, they gather more weapons and attract more fighters. The group is a direct offshoot of Al Qaeda in Iraq, Iraqi officials and former Iraqi insurgents say, which has contributed veteran fighters and weapons. "This is just a simple way of returning the favor to our Syrian brothers that fought with us on the lands of Iraq," said a veteran of Al Qaeda in Iraq, who said he helped lead the Nusra Front's efforts in Syria. The United States, sensing that time may be running out for Syria president Bashar al-Assad, hopes to isolate the group to prevent it from inheriting Syria, or fighting on after Mr. Assad's fall to pursue its goal of an Islamic state. As the United States pushes the Syrian opposition to organize a viable alternative government, it plans to blacklist the Nusra Front as a terrorist organization, making it illegal for Americans to have financial dealings with the group and prompting similar sanctions from Europe.

http://www.nytimes.com/2012/12/09/world/middleeast/syrian-rebels-tied-to-al-qaeda-play-key-role-in-war.html

Results

- 1069 extracted triples
  - 727 included in the merged ontology (non-redundant)
  - 524 in the reference ontology (cleaned up)
- And the winners are ...

Topic extraction

QUESTION:
Can we extract the *topics* of a text (often from a shortlist) from that text?

MATERIALS/SYSTEMS:
Alchemy, Open Calais, PoolParty KE
Informal output, e.g. "Culture & Politics, "Al_Qaeda"

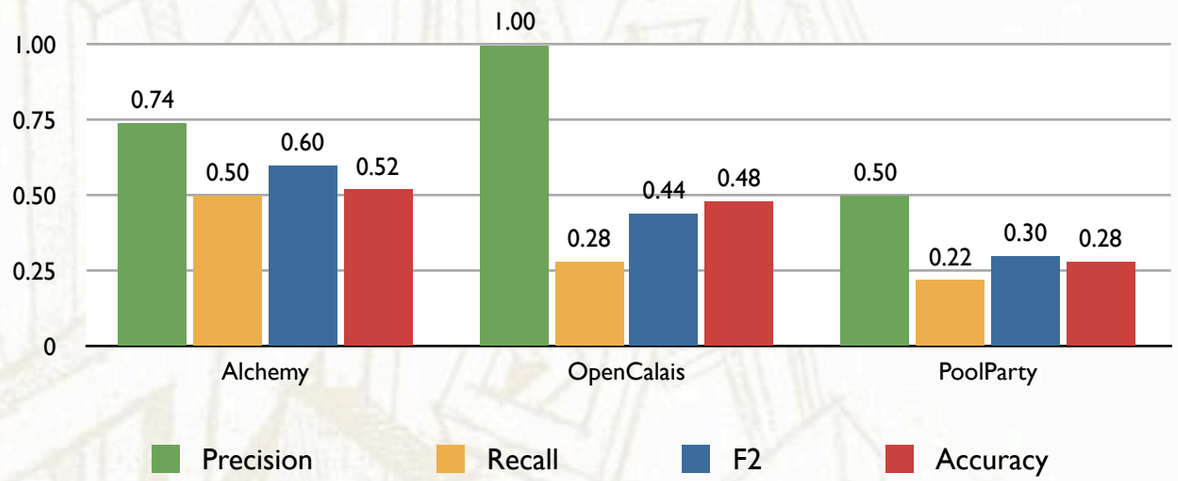
PROCEDURE:
Count and evaluate the tags, and generate `dc:subject` triples.

CONCLUSION:
Summing all extracted constructs, e.g.:

```
nusrafront:Text_NYT dc:subject
alchemy:Culture&Politics .
```

Alchemy results a little better than OpenCalais, which has better precision though.
For this task, we have accepted many sorts of topics: broad topics from a *given list*, topics resolved into *Wikipedia categories*, as well as *subjects* or *social tags*.

MEASURES	ALCHEMY	OPENCALAIS	POOLPARTY
PRECISION	0.74	1.00	0.50
RECALL	0.50	0.28	0.22
F2	0.60	0.44	0.30
ACCURACY	0.52	0.48	0.28



Topic Extraction	*Merged*	Alchemy	Open Calais	PoolParty KE	*Average*	*STD*
TP	*46*	23	13	10		
FP	*18*	8	0	10		
TP+FP	*64*	31	13	20		
FN		23	33	36		
TN		10	18	8		
P	*0.72*	0.74	**1.00**	0.50	*0.75*	*0.20*
R		**0.50**	0.28	0.22	*0.33*	*0.12*
F		**0.60**	0.44	0.30	*0.45*	*0.12*
A		**0.52**	**0.48**	0.28	*0.43*	*0.10*

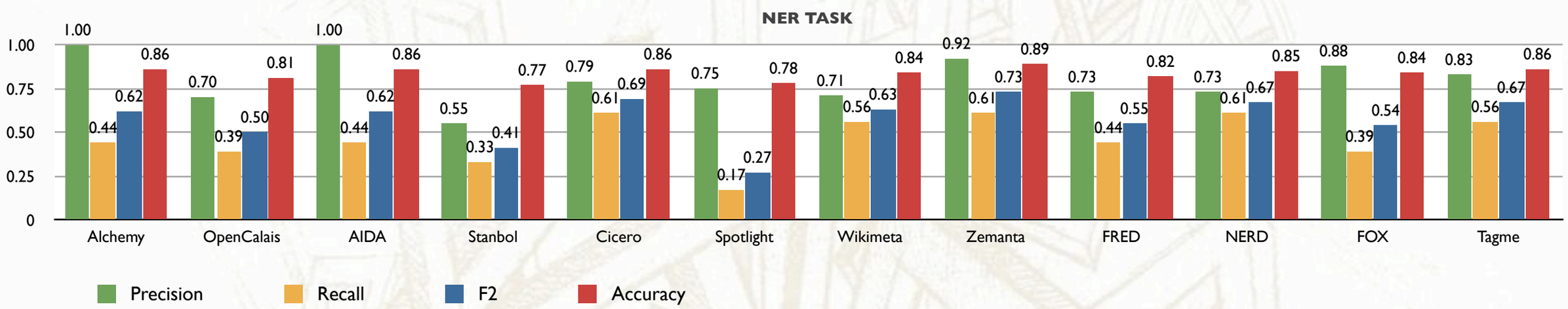
Named entity recognition

QUESTION:
Can we recognize individuals from text, e.g. "*The United States pushes the Syrian opposition*", or "*A veteran of Al Qaeda in Iraq*"?

MATERIALS/SYSTEMS:
Alchemy, Open Calais, AIDA, Stanbol, Cicero, Spotlight, Wikimeta, Zemanta, FRED, NERD, FOX, Tagme
RDF triples or informal output

PROCEDURE:
Count and evaluate recognized individuals.

CONCLUSION:
Zemanta results the best tool here, closely followed by Aida, Alchemy, Cicero and Tagme. The results confirm that the solutions to this task are mature.



Named Entity Recognition	*Merged*	Alchemy	Open Calais	AIDA	Stanbol	Cicero	Spotlight	Wikimeta	Zemanta	FRED	NERD	FOX	Tagme	*Average*	*STD*
TP	*18*	8	7	8	6	11	3	10	11	8	11	7	10		
FP	*55*	0	3	0	5	3	1	4	1	3	4	1	2		
TP+FP	*77*	8	7	8	10	14	4	14	12	11	15	8	12		
FN		10	11	10	12	7	15	8	7	10	7	11	8		
TN		55	52	55	50	52	54	51	54	52	51	54	53		
P	*0.25*	**1.00**	0.70	**1.00**	0.55	0.79	0.75	0.71	**0.92**	0.73	0.73	**0.88**	0.83	*0.80*	*0.13*
R		0.44	0.39	0.44	0.33	0.61	0.17	0.56	0.61	0.44	0.61	0.39	0.56	*0.46*	*0.13*
F		0.62	0.50	0.62	0.41	**0.69**	0.27	**0.63**	**0.73**	0.55	**0.67**	0.54	**0.67**	*0.57*	*0.12*
A		**0.86**	0.81	**0.86**	0.77	**0.86**	0.78	**0.84**	**0.89**	0.82	**0.85**	**0.84**	**0.86**	*0.84*	*0.04*

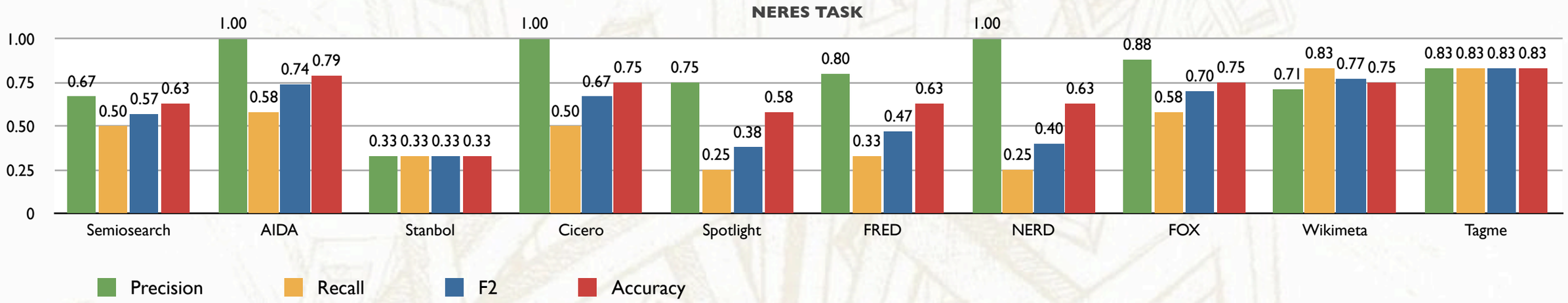
Named entity resolution

QUESTION:
Can we resolve the identity of recognized individuals from text, e.g. "*The United States pushes the Syrian opposition*", or "*A veteran of Al Qaeda in Iraq*"?

MATERIALS/SYSTEMS:
Semiosearch, AIDA, Stanbol, Cicero, Spotlight, Wikimeta, FRED, NERD, FOX, Tagme
RDF triples or informal output. Most resolutions are made on DBpedia/Wikipedia.

PROCEDURE:
Count and evaluate resolved individuals.

CONCLUSION:
Tagme results the best tool here, followed by Aida, Cicero and FOX.



Named Entity Resolution	*Merged*	Semiosearch	AIDA	Stanbol	Cicero	Spotlight	FRED	NERD	FOX	Wikimeta	Tagme	*Average*	*STD*
TP	*12*	6	7	4	6	3	4	3	7	10	10		
FP	*12*	3	0	8	0	1	1	0	1	4	2		
TP+FP	*24*	9	7	12	6	4	5	3	8	14	14		
FN		6	5	8	6	9	8	9	5	2	2		
TN		9	12	4	12	11	11	12	11	8	10		
P	*0.50*	0.67	**1.00**	0.33	**1.00**	0.75	0.80	**1.00**	0.88	0.71	0.83	*0.80*	*0.20*
R		0.50	0.58	0.33	0.50	0.25	0.33	0.25	0.58	0.83	0.83	*0.50*	*0.20*
F		0.57	**0.74**	0.33	**0.67**	0.38	0.47	0.40	**0.70**	**0.77**	**0.83**	*0.59*	*0.17*
A		0.63	**0.79**	0.33	**0.75**	0.58	0.63	0.63	**0.75**	**0.75**	**0.83**	*0.67*	*0.14*

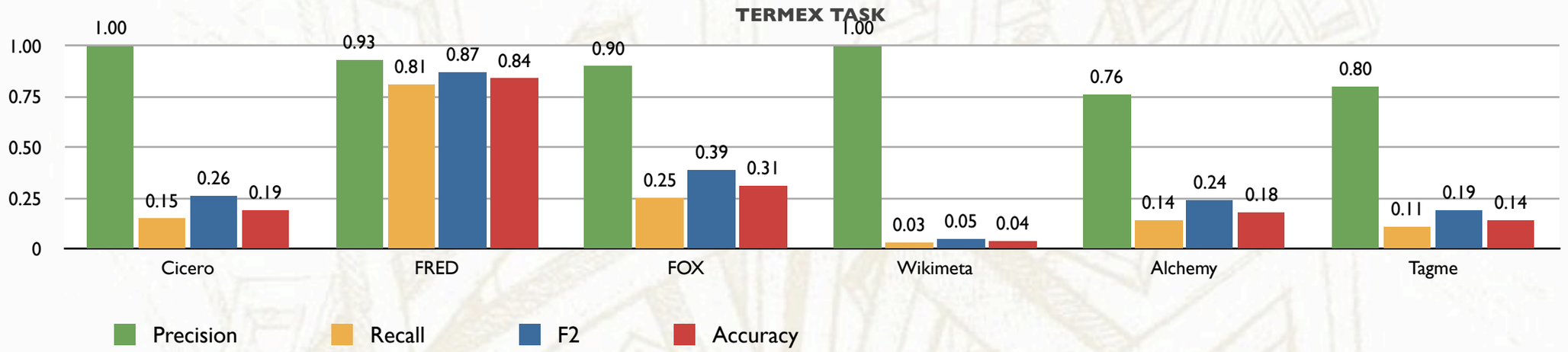
Terminology extraction

QUESTION:
Can we recognize terms/concepts from text, e.g. "*Its fighters, a small minority of the rebels*"?

MATERIALS/SYSTEMS:
Alchemy, Cicero, Wikimeta, FRED, FOX, Tagme
RDF triples or informal output

PROCEDURE:
Count and evaluate extracted terms/concepts.

CONCLUSION:
FRED results the best tool here, suggesting that deep parsing and logical processing might be better than purely statistical approaches.
Consider that none of the tools analyzed is specifically a "term extractor", but the extraction is a byproduct of other functionalities, e.g. NER or keyword extraction.
As a matter of fact, only some tools (Alchemy, FRED) explicitly distinguish between extracted named entities and extracted terms/concepts.
For comparison, a term extractor like TerMine has results very similar to Alchemy.



Term (concept) Extration	*Merged*	Cicero	FRED	FOX	Wikimeta	Alchemy	Tagme	*Average*	*STD*
TP	*112*	17	91	28	3	16	12		
FP	*10*	0	7	3	0	5	3		
TP+FP	*122*	17	98	31	3	21	15		
FN		95	21	84	109	96	100		
TN		10	3	7	10	5	7		
P	*0.92*	**1.00**	0.93	0.90	**1.00**	0.76	0.80	*0.90*	*0.09*
R		0.15	0.81	0.25	0.03	0.14	0.11	*0.25*	*0.26*
F		0.26	**0.87**	0.39	0.05	0.24	0.19	*0.33*	*0.26*
A		0.19	**0.84**	0.31	0.04	0.18	0.14	*0.28*	*0.26*

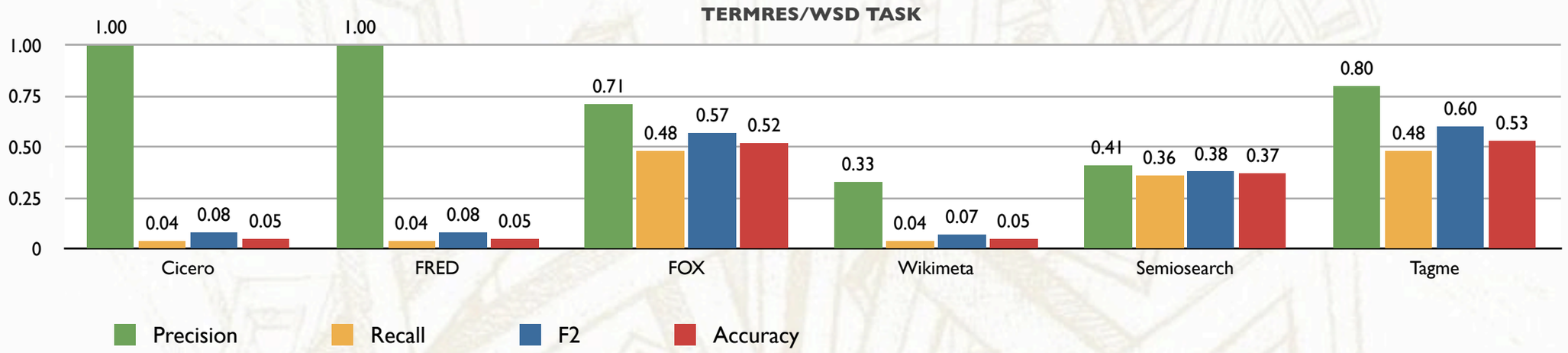
Word sense disambiguation

QUESTION:
Can we resolve the identity of concepts, i.e. can we disambiguate the sense of a word/term from text, e.g. "*Its fighters, a small minority of the rebels*"?

MATERIALS/SYSTEMS:
Cicero, Wikimeta, FRED, FOX, Tagme, Semiosearch
RDF triples or informal output. Most resolutions are made on DBpedia/Wikipedia.

PROCEDURE:
Count and evaluate resolved terms/concepts.

CONCLUSION:
Tagme results the best tool here, closely followed by FOX. Tagme seems then the best on resolution tasks.
Consider that none of the tools analyzed is specifically a "word sense disambiguator", but the term resolution is a byproduct of named entity resolution.
By comparison, a top-class WSD tool like IMS disambiguates most of the words (112) against WordNet, typically with precision around .7-.8.



Term (concept) Resolution	*Merged*	Cicero	FRED	FOX	Wikimeta	Semiosearch	Tagme	*Average*	*STD*
TP	*25*	1	1	12	1	9	12		
FP	*20*	0	0	5	2	13	3		
TP+FP	*45*	1	1	17	3	22	15		
FN		24	24	13	24	16	13		
TN		20	20	15	18	7	17		
P	*0.56*	**1.00**	**1.00**	0.71	0.33	0.41	0.80	*0.71*	*0.26*
R		0.04	0.04	0.48	0.04	0.36	0.48	*0.24*	*0.20*
F		0.08	0.08	**0.57**	0.07	0.38	**0.60**	*0.30*	*0.23*
A		0.05	0.05	**0.52**	0.05	**0.37**	**0.53**	*0.26*	*0.22*

Sense tagging (type induction)

QUESTION:

Can we induce types of individuals (typically using a small ontology) from text even if the type is not mentioned, e.g. "*The United States pushes the Syrian opposition*", or "*A veteran of Al Qaeda in Iraq*"?

MATERIALS/SYSTEMS:

Alchemy, Open Calais, AIDA, Stanbol, Cicero, Spotlight, Wikimeta, Zemanta, FRED, NERD, FOX
RDF triples or TSV output
`rdf:type` assertions
Informal formatted output, e.g.
Syrian American(org)

PROCEDURE:

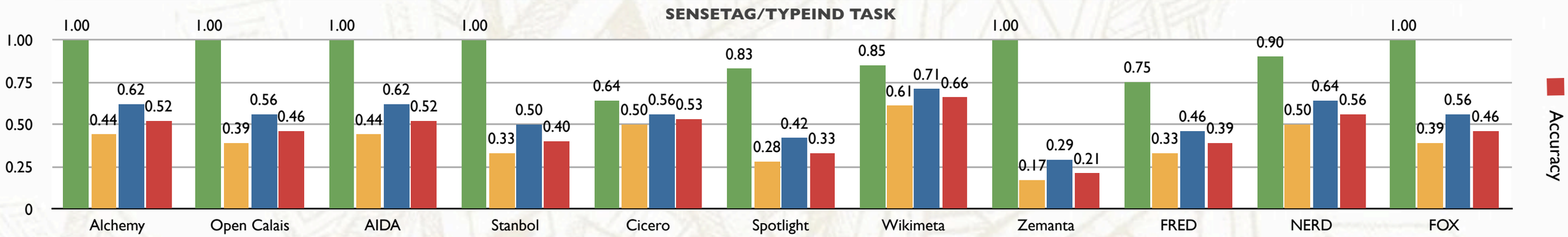
Count and evaluate `rdf:type` assertions, as well as informal tags, as induced types.

CONCLUSION:

Summing all extracted constructs, e.g.:

```
nusrafront:Syria rdf:type wikimeta:loc .
nusrafront:Nusra_front rdf:type foaf:Organisation .
```

Wikimeta results the best tool here, even though a full precision is achieved by most other tools.



Sense Tagging (type induction)	*Merged*	Alchemy	Open Calais	AIDA	Stanbol	Cicero	Spotlight	Wikimeta	Zemanta	FRED	NERD	FOX	*Average*	*STD*
TP	*14*	8	7	8	6	9	5	11	3	6	9	7		
FP	*5*	0	0	0	0	5	1	2	0	2	1	0		
TP+FP	*19*	8	7	8	6	14	6	13	3	8	10	7		
FN		6	7	6	8	5	9	3	11	8	5	7		
TN		5	5	5	5	0	4	3	5	3	4	5		
P	*0.74*	**1.00**	**1.00**	**1.00**	**1.00**	0.64	0.83	0.85	**1.00**	0.75	0.90	**1.00**	*0.91*	*0.12*
R		0.57	0.50	0.57	0.43	0.64	0.36	0.79	0.21	0.43	0.64	0.50	*0.51*	*0.12*
F		**0.73**	0.67	**0.73**	0.60	0.64	0.50	**0.81**	0.35	0.55	**0.75**	0.67	*0.64*	*0.11*
A		**0.64**	0.57	**0.64**	0.50	**0.64**	0.42	**0.80**	0.27	0.48	**0.69**	0.57	*0.57*	*0.12*

Relation extraction

QUESTION:
Can we extract binary relationships from a text, e.g. *Money flows to the group*?

MATERIALS/APPARATUS:
Alchemy, Cicero, FRED, reVerb, Ollie
RDF triples or TSV output
`owl:ObjectProperty` assertions
`dul:Event` entities
We have excluded semantic role relationships, counted in the next tasks.
Informal formatting of TSV output, e.g.
Money @ flows to
@ the group

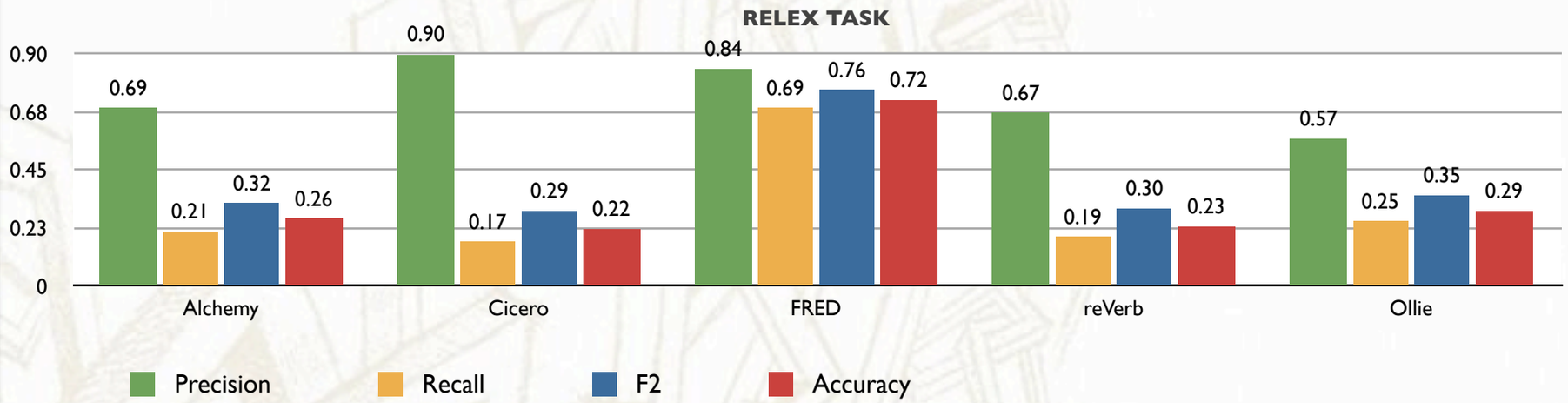
PROCEDURE:
Count and evaluate `owl:ObjectProperty` assertions, as well as formatted middle terms in TSV, as property assertions.

CONCLUSION:
Summing all extracted constructs, e.g.:

```
nusrafront:Money flows_to the_group .
nusrafront:approval_1 nusrafront:from
nusrafront:Al_qaeda .
```

FRED results the best tool here, while Cicero is best for precision.
Probably a sign of the typical tradeoff between logic-oriented (deep) and stochastically-oriented (shallow) parsing.

MEASURES	ALCHEMY	CICERO	FRED	REVERB	OLLIE
PRECISION	0.69	0.90	0.84	0.67	0.57
RECALL	0.21	0.17	0.69	0.19	0.25
F2	0.32	0.29	0.76	0.30	0.35
ACCURACY	0.26	0.22	0.72	0.23	0.29



(Binary) Relation(ship) Extraction	*Merged*	Alchemy	Cicero	FRED	reVerb	Ollie	*Average*	*STD*
TP	*52*	11	9	36	10	13		
FP	*27*	5	1	7	5	10		
TP+FP	*79*	16	10	43	15	23		
FN		41	43	16	42	39		
TN		22	26	20	22	17		
P	*0.66*	0.69	**0.90**	**0.84**	0.67	0.57	*0.77*	*0.10*
R		0.21	0.17	0.69	0.19	0.25	*0.32*	*0.22*
F		0.32	0.29	**0.76**	0.30	0.35	*0.42*	*0.20*
A		0.26	0.22	**0.72**	0.23	0.29	*0.36*	*0.21*

Event extraction

QUESTION:
Can we extract events from a text, e.g. *This is just a simple way of returning the favor to our Syrian brothers*?

MATERIALS/APPARATUS:
FRED, Cicero, Open Calais
RDF triples or textual output
`owl:ObjectProperty` assertions
`dul:Event` entities
Informal formatting of textual output

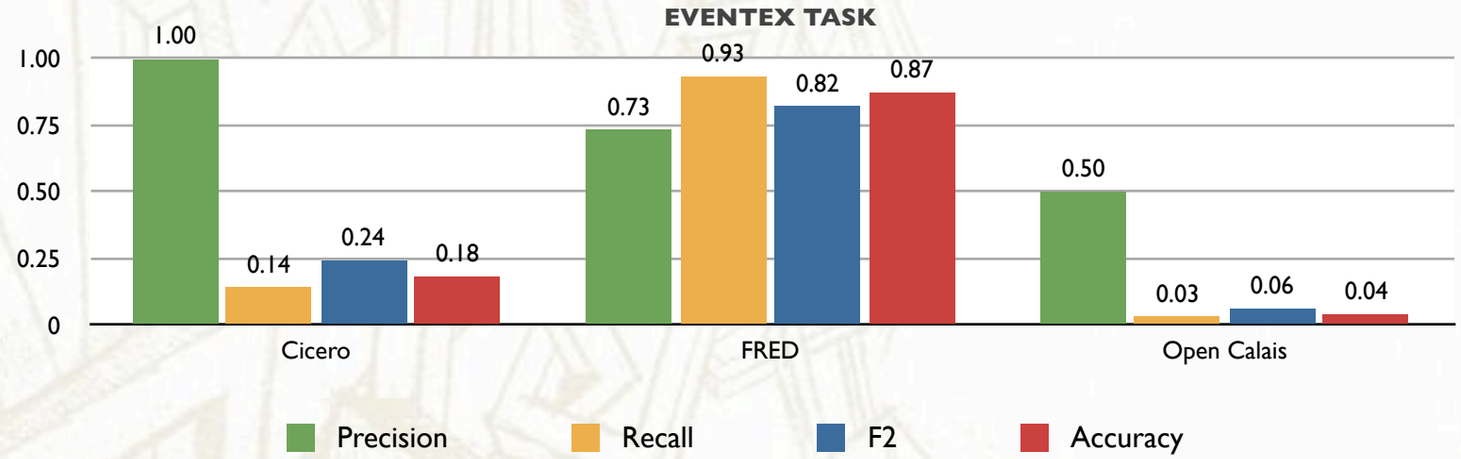
PROCEDURE:
Count and evaluate `owl:ObjectProperty` assertions with `dul:Event` as subject, as well as formatted middle terms, as labeled semantic roles.

CONCLUSION:
Summing all extracted constructs, e.g.:

```
nusrafront:return_1
                verbnet:Agent nusrafront:way_1 ;
                verbnet:Recipient nusrafront:brother_1 ;
                verbnet:Theme nusrafront:favor_1 .
```

FRED results the best tool here, except for precision.

MEASURES	CICERO	FRED	OPEN CALAIS
PRECISION	1.00	0.73	0.50
RECALL	0.14	0.93	0.03
F2	0.24	0.82	0.06
ACCURACY	0.18	0.87	0.04



Event (n-ary relationship) Extraction	*Merged*	Cicero	FRED	Open Calais	*Average*	*STD*
TP	*29*	4	27	1		
FP	*11*	0	10	1		
TP+FP	*40*	4	37	2		
FN		25	2	28		
TN		11	1	10		
P	*0.73*	**1.00**	0.73	0.50	*0.74*	*0.20*
R		0.14	0.93	0.03	*0.37*	*0.40*
F		0.24	**0.82**	0.06	*0.38*	*0.32*
A		0.18	**0.87**	0.04	*0.36*	*0.36*

SRL and frame detection

QUESTION:

Can we extract roles and frames from a text, e.g. *This is just a simple way of returning the favor to our Syrian brothers*?

MATERIALS/APPARATUS:

FRED, Cicero
RDF triples or textual output
`owl:ObjectProperty` assertions
`dul:Event` entities
Informal formatting of textual output

PROCEDURE:

Count and evaluate `owl:ObjectProperty` assertions with `dul:Event` as subject, as well as formatted middle terms, as labeled semantic roles.
Count and evaluate `dul:Event` *types* and formatted left terms as detected frames.
Count and evaluate `owl:ObjectProperty` assertions objects w.r.t *types* of fillers of the semantic role.
Count and evaluate coreference resolution cases.

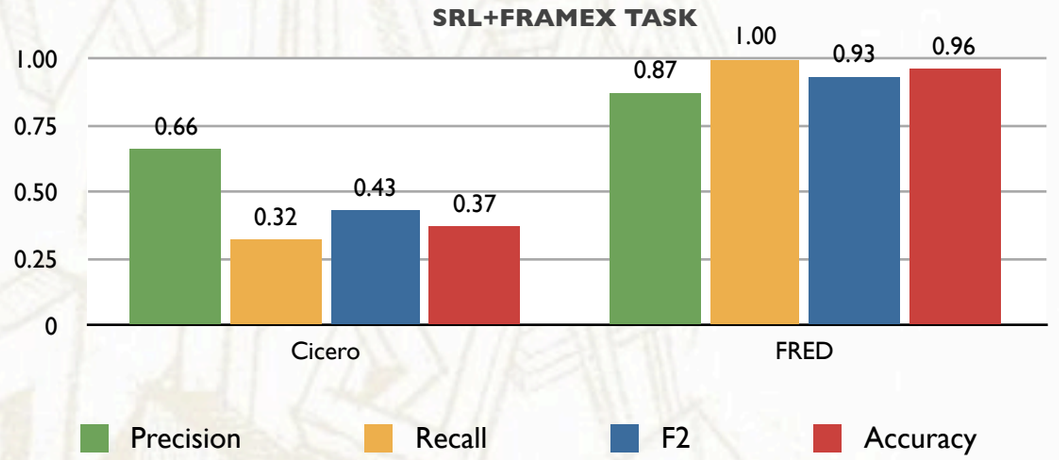
CONCLUSION:

Summing all extracted constructs, e.g.:

```
verbnet:Agent rdfs:domain nusrafront:Return ;
verbnet:Agent rdfs:range nusrafront:Way ;
verbnet:Recipient rdfs:domain nusrafront:Return ;
verbnet:Recipient rdfs:range nusrafront:Brother ;
verbnet:Theme rdfs:domain nusrafront:Return ;
verbnet:Theme rdfs:range nusrafront:Favor .
```

FRED results the best tool here on all measures.

MEASURES	CICERO	FRED
PRECISION	0.66	0.87
RECALL	0.32	1.00
F2	0.43	0.93
ACCURACY	0.37	0.96



Semantic Role Labelling + Frame Detection (All elements)	*Merged*	*Gold*	Cicero	FRED	*Average*	*STD*
TP	*225*	*288*	73	225		
FP	*72*		38	35		
TP+FP	*297*		111	260		
FN			152	0		
TN			34	37		
P	*0.76*		0.66	**0.87**	*0.76*	*0.10*
R			0.32	1.00	*0.66*	*0.26*
F			0.43	**0.93**	*0.68*	*0.23*
A			0.37	**0.96**	*0.67*	*0.36*

Extraction from compositional lexicon

QUESTION:
Can we extract relationships from compositional lexicon in a text, e.g. *oil field*?

MATERIALS/APPARATUS:
FRED
RDF triples
`rdfs:subClassOf` relation
`dul:associatedWith` relation

PROCEDURE:
Count `rdfs:subClassOf` triples as induced taxonomic relations.
Count `dul:associatedWith` triples as induced generic relations.

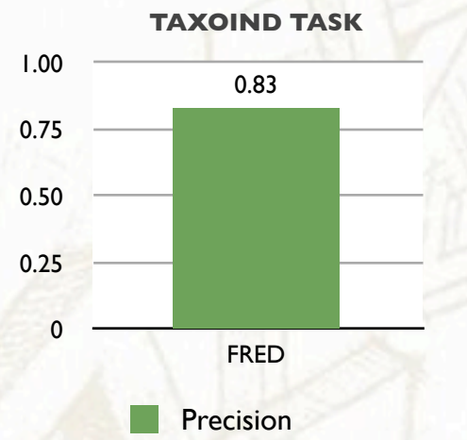
CONCLUSION:
FRED extracts 24 taxonomic relationships (.83 precision), e.g.:

```
OilField rdfs:subClassOf Field
```

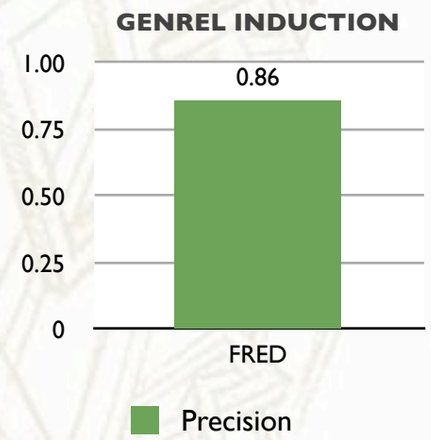
and 6 generic relationships (. 86 precision), e.g.:

```
oil_1 dul:associatedWith field_1
```

MEASURES	FRED
PRECISION	0.83



MEASURES	FRED
PRECISION	0.86



Taxonomy (subclass axiom) Induction	FRED
TP	20
FP	4
TP+FP	24
FN	
TN	
P	0.83
R	n/a
F	n/a

Generic relation Induction	FRED
TP	6
FP	1
TP+FP	7
FN	
TN	
P	0.86
R	n/a
F	n/a

Conclusions

- Feasibility of a comparison among KE/machine reading tools when used for SW tasks
- Tools are not homogeneous in semantic assumptions, data structures, text span tagging, settings, output, availability, licensing, performance time, etc.
- Information measures greatly vary, except for NER
- NER and sense tagging seem better supported by purely statistical algorithms, relation/event extraction and compositional semantics seem better supported by deep parsing-based tools
- Integration of different tools seems the natural way to go
  - But interoperability protocols/platforms should support explicit declaration of semantic correspondences, and customization (Stanbol Rules on the right track here)
  - Benchmarking is key to filter the worst output of selected tools!

More parameters and measures (1/2)

- System-level features
  - time performance, export capabilities, standard compliance
    - no evaluation of time performance: tested conditions were heterogeneous
    - some big gaps have been detected, please ask personally
  - export and standard are generally low, and are not focused in this analysis

More parameters and measures (2/2)

- Structural measures of the produced ontologies
  - axioms, density, presence of linguistic annotations, textual grounding, etc.
  - generally low except for FRED and somehow Cicero, not focused here

More tasks?

- There are more tasks that are not really addressed by NLP
  - schema-level logical structure extraction
    - class equivalence, union of classes, class covering, class disjointness, ...
      - but: cf. automatic formalization of glosses, discourse analysis, DRT, Markov Logic
  - data-level logical structure extraction
    - entity linking, individual conjunction, individual disjunction, fact negation, factual entailment, ...

Ground truth

- Hard to establish, due to many possible tasks when evaluating the extracted knowledge
  - A workaround has been adopted
- *Completeness*
  - Following IR practices, each tool contributes its results to a merged ontology that is used as the reference result set
    - cf. retrieval evaluation with incomplete information: supervised results from different methods are merged
- *Correctness*
  - Cleanup of the merged ontology, in order to produce a reference ontology
    - Cleanup performed by 3 ontology design experts that read the text, and exclude intuition-clashing results (simpler than deciding on a unique reference ontology/gold standard)

Thanks!

Demos available tomorrow night for FRED (and Tìpalo)