NERD: A Framework for Evaluating Named Entity Recognition Tools in the Web of Data

EURECOM
Sophia Antipolis

OBJECTIVE

NERD is an evaluation framework which records and analyzes ratings of Named Entity (NE) extraction and disambiguation tools, working on English plain text articles, performed by human beings. NERD enables the comparison of different entity extractors which expose APIs such as AlchemyAPI, DBPedia Spotlight, Evri, Extractiv, OpenCalais, Yahoo! Term extraction and Zemanta. Given an article and a particular tool, a user can assess the precision of the named entities extracted, their typing and linked data URI provided for disambiguation and their subjective relevance for the text. All user interactions are stored in a database.
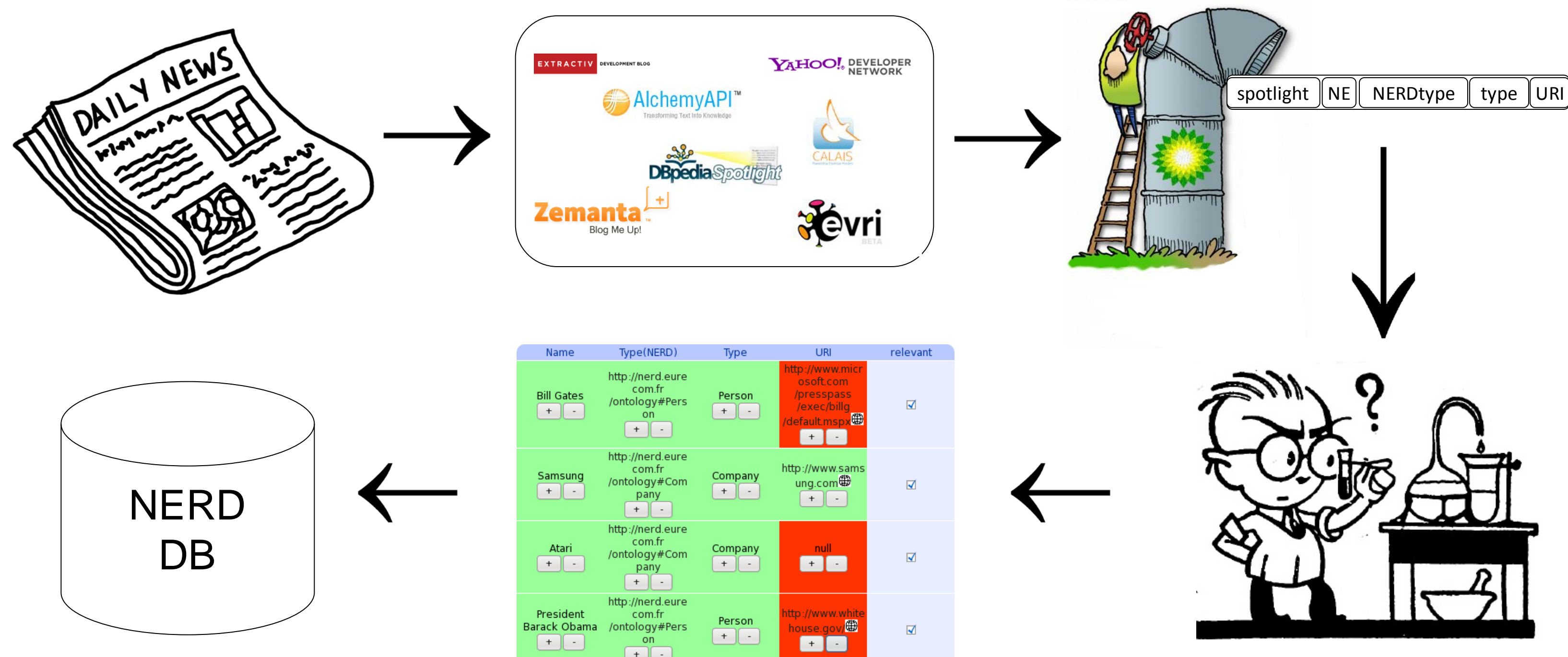
EVALUATION

What is a Named Entity?

A *named entity* is a name of a person or an organization, a location, a brand, a product, a numeric expression including time, date, money and percent found in a sentence [1].
NERD unifies the outputs provide by the supported extractors and, keeping the provenance information (name of the extractor), it shows an output tuple t:

| Named Entity | NERDType | Type | disambiguated URI |

NERD classifies each Named Entity also according the NERD ontology and visualizes the value in the NERDType field.

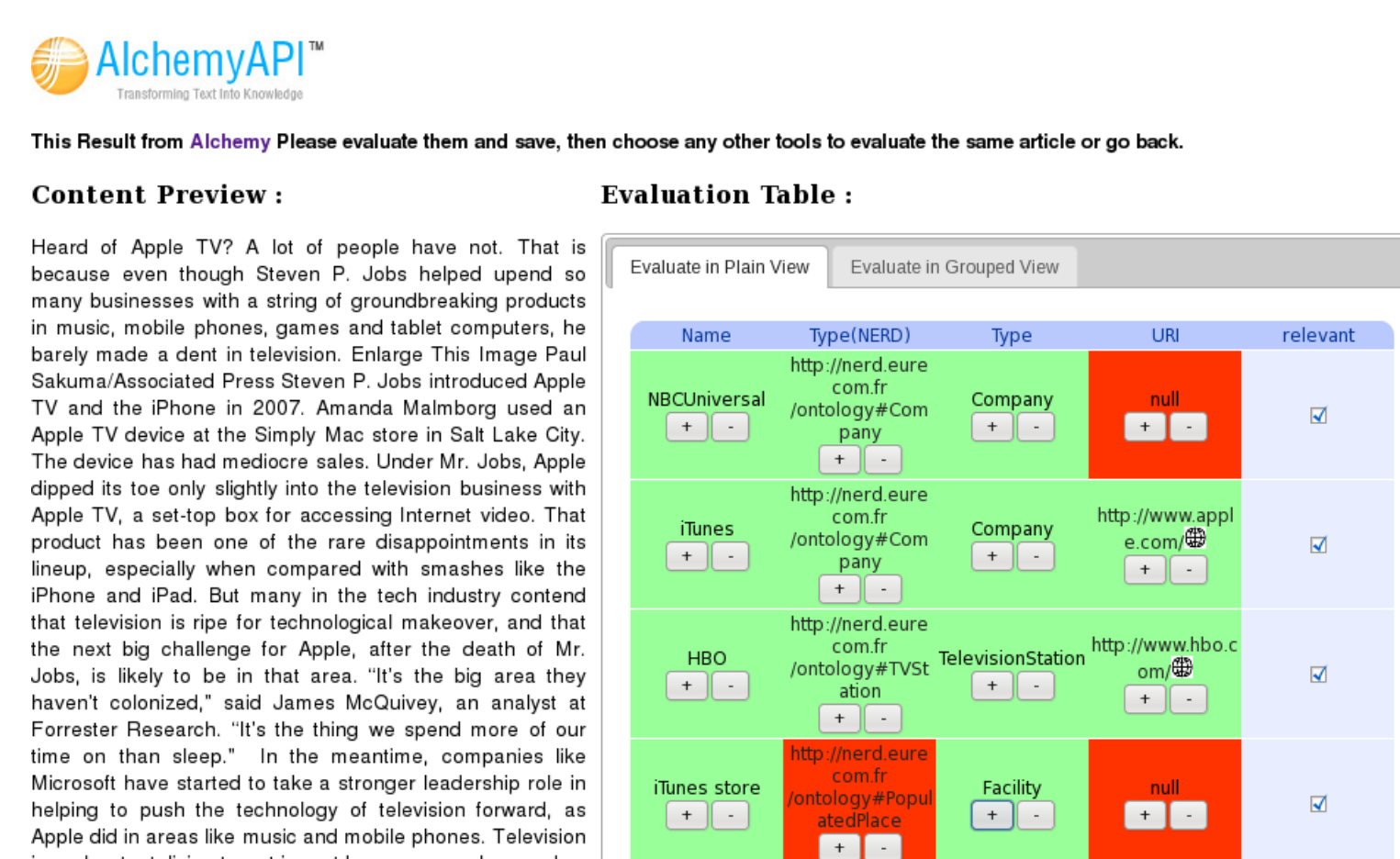### What does the evaluation consist of ?

Human beings rate the output tuple, giving for each field a Boolean score. In addition, the evaluator judges the relevance of the pair NE and type if it is actually relevant for the text being analyzed. In the case where no type or no disambiguation URI was provided by the tool, it would be considered as false.

NERD

NERD REST API[1]

/api/service
/api/doc
/api/annotation/{extractor}
/api/extraction
/api/evaluation

NERD UI[2]

NERD ontology[3]

One of the differences among these NE extractors is the taxonomy/ontology used to classify entity types. We develop the NERD ontology by manually aligning the different classes using their definitions and providing a best coverage of the principal axioms.
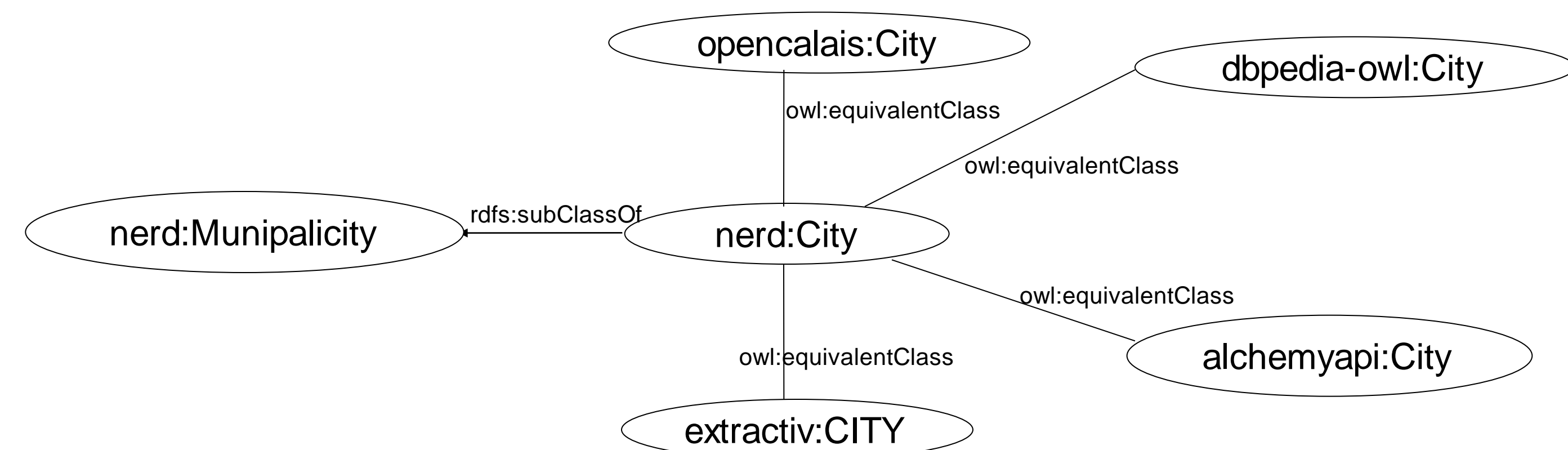
GOLDEN DATASET

In [2] we proposed a golden-set where 4 human assessors rated 10 different articles (BBC and NYT) with 5 NE extractors whose APIs are publicly available.

	AlchemyAPI	DBpedia Spotlight	Extractiv	OpenCalais	Zemanta
NE	slight	poor	slight	fair	slight
type	poor	fair	substantial	moderate	poor
URI	substantial	poor	almost	almost	fair
relevant	slight	poor	slight	fair	fair

We explained the low agreement among the 4 raters mainly due to the length of the text to analyze. In this conference, we are inviting all attendants to evaluate a subset of [3] for all the extractors and to create a new golden dataset. We also report the precision and relevance scores.

	AlchemyAPI	DBpedia Spotlight	Extractiv	OpenCalais	Zemanta
overall precision	0.7054	0.4915	0.611	0.5396	0.6463
relevant score	0.9005	0.5525	0.6805	0.8224	0.8800

We invite to create a new golden dataset!

• Science: "Google Cars Drive Themselves", **http://nyti.ms/9p19i8**
"*MOUNTAIN VIEW, Calif. — Anyone driving the twists of Highway 1 between San Francisco and Los Angeles recently may have glimpsed a Toyota Prius with a curious funnel-like cylinder on the roof. Harder to notice was that the person at the wheel was not actually driving.*"

• Art: "Chaos and Classicism, Interwar Art, at Guggenheim", **http://nyti.ms/aXO2qm**
"*Boilerplate is safe box office, and we've gotten our share lately. So it's great that the Guggenheim Museum is giving us the opposite in its major fall exhibition, "Chaos and Classicism: Art in France, Italy, and Germany, 1918-1936."*"

REFERENCES

[1] Ralph Grishman and Beth Sundheim. Message Understanding Conference-6: a brief history. In 16th International Conference on Computational linguistics (COLING'96), pages 466–471, Copenhagen, Denmark, 1996.
[2] Giuseppe Rizzo and Raphaël Troncy. NERD: Evaluating Named Entity Recognition Tools in the Web of Data. In (ISWC'11) Workshop on Web Scale Knowledge Extraction (WEKEX'11), Bonn, Germany, October 23-27, 2011.
[3] Pablo Mendes, Max Jakob, Andrés García-Silva and Christian Bizer. DBpedia Spotlight: Shedding Light on the Web of Documents. In the Proceedings of the 7th International Conference on Semantic Systems (I-Semantics). Graz, Austria, 7–9 September 2011.

ACKNOWLEDGMENT

CONTACTS

Giuseppe Rizzo — **giuseppe.rizzo@eurecom.fr**
Raphaël Troncy — **raphael.troncy@eurecom.fr**