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.

NERD ontology

One of the differences among these NE extractors is the taxonomy/ontology used to classify entity types. We develop the NERD ontology3

NERD REST API

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

NERD UI

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:

<table>
<thead>
<tr>
<th>Named Entity</th>
<th>NERDType</th>
<th>Type</th>
<th>disambiguated URI</th>
</tr>
</thead>
</table>

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.

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.

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.

We invite to create a new golden dataset!

• Science: “Google Cars Drive Themselves”, http://nyt.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, Intervar Art, at Guggenheim”, http://nyt.ms/71O2am
“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


CONTACTS

Giuseppe Rizzo
raphael.troncy@eurecom.fr

ACKNOWLEDGMENT

This work was supported by the French Ministry of Industry (Innovative Web call) under contract 09.2.93.0966, “Collaborative Annotation for Video Accessibility” (ACAV).

CONTACTS

Giuseppe Rizzo
raphael.troncy@eurecom.fr