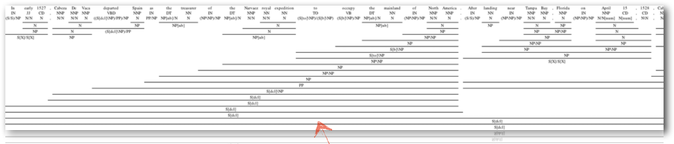
A Machine Reader for the Semantic Web

Aldo Gangemi[1,2], Valentina Presutti[1], Francesco Draicchio[1], Andrea Giovanni Nuzzolese[1], Diego Reforgiato Recupero[1]

1-STLab, ISTC-CNR, Rome, 2-LIPN, Un. Paris13-CNRS-Sorbonne Cité

{aldo.gangemi,valentina.presutti}@cnr.it

http://stlab.istc.cnr.it



Abstract

We have implemented a novel approach for robust ontology and linked data production from natural language texts by combining Discourse Representation Theory (DRT), linguistic frame semantics, and ontology design patterns.

We have defined a Semantic Web-oriented translation from DRT and lexical semantics to RDF/OWL for *machine reading* on the SW.

We have designed FRED, a tool for converting text into internally well-connected and linked-data-ready ontologies in web-service-acceptable time.

What?

The problem: accurate knowledge extraction from text, aka *deep machine reading in RDF and OWL*.

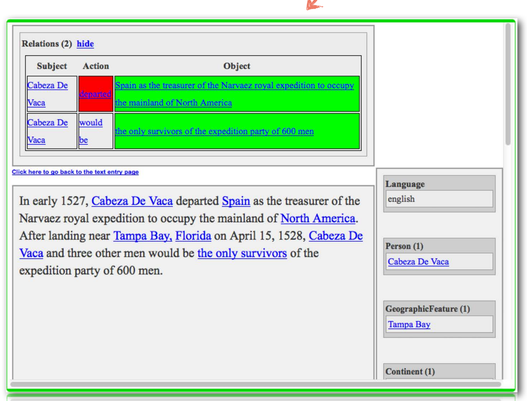
The SOA: shallow knowledge extraction [2] – NER, sense tagging, binary relations ...

FRED's machine reading solution [5]: RDF-OWL from categorial deep parsing, Boxer [1] computational semantics representation as DRT [4], frame semantics, ontology design patterns-based heuristics, enriched with entity resolution and linking, word sense disambiguation and vocabulary mapping.

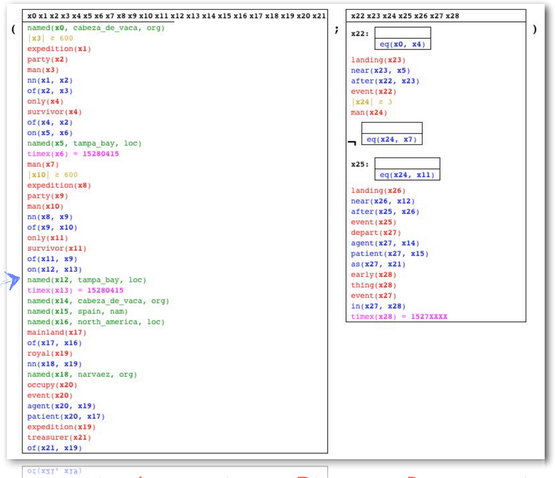
Example

"*Cabeza de Vaca departed Spain in early 1527 as the treasurer of the Narvaez royal expedition to occupy the mainland of North America. After landing near Tampa Bay, Florida on April 15, 1528, Cabeza De Vaca and three other men would be the only survivors of the expedition party of 600 men.*"

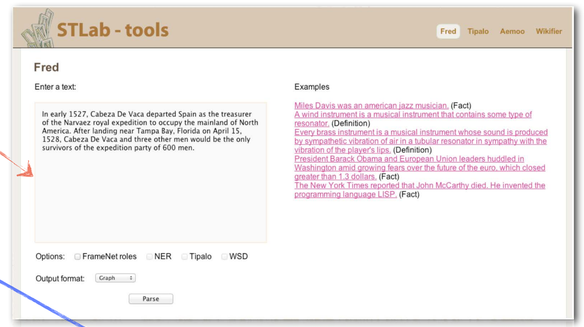
Shallow Parsing (e.g. Alchemy, http://www.alchemyapi.com/) vs. FRED's Machine Reading.

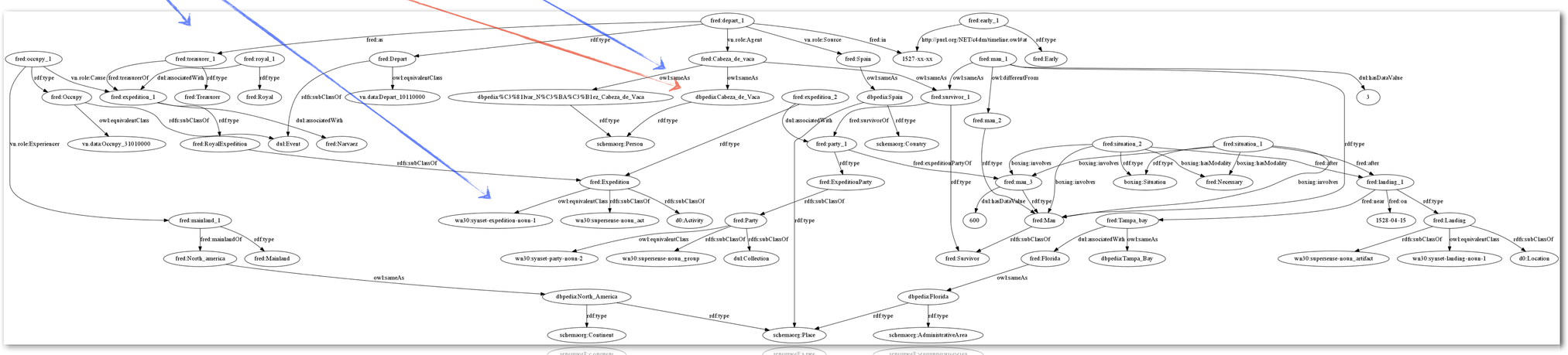


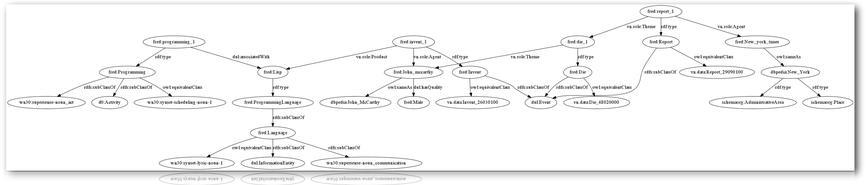
How?



Deep parsing and computational semantics to Discourse Representation Structures (DRS "boxing" style), plus many sorts of translations and linking.







FRED is then a machine reader for the Semantic Web.

FRED uses a lot of heuristics for refactoring DRS and NLP output into LOD-ready graphs, improving their design, and enriching/linking them to LOD and common vocabularies (FOAF, SKOS, DBpedia, schema.org, DOLCE+DnS, Earmark – NIF and Lemon forthcoming).

FRED uses Apache Stanbol for NER and resolution to DBpedia, IMS/UKB for WSD to WordNet and VerbNet.

New FRED release produces Earmark text spans and its API provides several RDF serializations.

It addresses an extended set of logical constructs, including *negation, modalities, situations,* and *tense representation*:

FRED: http://wit.istc.cnr.it/stlab-tools/fred

Tìpalo [3] is an entity type induction tool based on FRED:

Tìpalo: http://wit.istc.cnr.it/stlab-tools/tipalo

Sèntilo is an opinion mining tool based on FRED:

Sèntilo: http://wit.istc.cnr.it/stlab-tools/sentilo

References

[1] J. Bos, Wide-Coverage Semantic Analysis with Boxer, Semantics in Text Processing (Johan Bos and Rodolfo Delmonte, eds.), College Publications, 2008, pp. 277–286.

[2] A. Gangemi. A Comparison of Knowledge Extraction Tools for the Semantic Web. Proceedings of ESWC2013, LNCS, Springer, 2013.

[3] A. Gangemi, A. G. Nuzzolese, V. Presutti, F. Draicchio, A. Musetti and P. Ciancarini. Automatic Typing of DBpedia Entities. Proceedings of ISWC2012, the Tenth International Semantic Web Conference, LNCS, Springer, 2012.

[4] H: Kamp, A theory of truth and semantic representation, Formal Methods in the Study of Language (Jeroen A. G. Groenendijk, Teo M. V. Janssen, and Martin B. J. Stokhof, eds.), vol. 1, Mathematisch Centrum, 1981, pp. 277–322.

[5] V. Presutti, F. Draicchio and A. Gangemi. Knowledge Extraction based on Discourse Representation Theory and Linguistic Frames. A. ten Teije and J. Völker (eds.): Proceedings of the Conference on Knowledge Engineering and Knowledge Management (EKAW2012), LNCS, Springer, 2012.