Semantic Web Machine Reading with FRED

A. Gangemi*, V. Presutti*, D. Reforgiato Recupero^, A.G. Nuzzolese^, F. Draicchio^, M. Mongiovi^

^STLab, ISTC-CNR (Rome, Italy)
*LIPN, Paris Nord University

ISWC2017, Journal Track
October 24th, 2017

Outline

• Overview of FRED
• Main mappings and heuristics (the method behind)
• FRED pipeline (implementation)
• Quality, importance, impact (evaluation)
• Conclusion
Outline

- Overview of FRED
- Main mappings and heuristics (the method behind)
- FRED pipeline (implementation)
- Quality, importance, impact (evaluation)
- Conclusion
FRED: SW machine reader

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

http://wit.istc.cnr.it/stlab-tools/fred/fredlib
Frame-based formal representation

“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”
Multiple NLP tools

Stanford’s dependency parsing, coreference resolution

(ROOT
(S
(NP (NNP John) (NNP Coltrane))
(VP (VBD played)
  (PP (IN with)
    (NP (NNP Miles) (NNP Davis)))
  (PP (IN in)
    (NP (NNP Kind))
    (PP (IN of)
      (NP (NNP Blue)))))))

Boxer’s DRT boxing

<table>
<thead>
<tr>
<th>x0</th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
</tr>
</thead>
<tbody>
<tr>
<td>root(ROOT-0, played-3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nn(Coltrane-2, John-1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nsubj(played-3, Coltrane-2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nn(Davis-6, Miles-5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prep_with(played-3, Davis-6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prep_in(played-3, Kind-8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prep_of(Kind-8, Blue-10)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>e4</th>
</tr>
</thead>
<tbody>
<tr>
<td>play(e4)</td>
</tr>
<tr>
<td>Actor1(e4,x0)</td>
</tr>
<tr>
<td>named(x0,john_coltrane,per)</td>
</tr>
<tr>
<td>named(x1,kind,loc)</td>
</tr>
<tr>
<td>named(x2,blue,loc)</td>
</tr>
<tr>
<td>named(x3,miles_davis,loc)</td>
</tr>
<tr>
<td>in(e4,x1)</td>
</tr>
<tr>
<td>Actor2(e4,x3)</td>
</tr>
</tbody>
</table>

Babelfy’s (or UKB’s) linking

Language recogniser +
BING off the shelf translator

Tagme’s linking
<table>
<thead>
<tr>
<th>Semantic web resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBpedia</td>
</tr>
<tr>
<td>schema.org</td>
</tr>
<tr>
<td>WordNet-RDF</td>
</tr>
<tr>
<td>BabelNet</td>
</tr>
<tr>
<td>VerbNet-RDF</td>
</tr>
<tr>
<td>FrameNet-OWL</td>
</tr>
<tr>
<td>OntoWordNet</td>
</tr>
<tr>
<td>DOLCE-Zero</td>
</tr>
<tr>
<td>Ontology Design Patterns</td>
</tr>
<tr>
<td>NIF</td>
</tr>
<tr>
<td>Earmark</td>
</tr>
</tbody>
</table>

Now everything formally integrated in Framester! [https://github.com/framester/Framester/wiki/Framester-Documentation](https://github.com/framester/Framester/wiki/Framester-Documentation)
Main contributions of FRED

• A novel form of machine reading: from text to knowledge graph

• A method for integrating multiple NLP task outputs into a knowledge graph
  • Formal mappings between different theories (implemented as a set of heuristics)
  • Ontology design patterns for the Semantic Web
  • Reference cognitive semantics for interpretation: Frame Semantics
Outline

- Overview of FRED
- Main mappings and heuristics (the method behind)
- FRED pipeline (implementation)
- Quality, importance, impact (evaluation)
- Conclusion
Main mappings and heuristics

- From DRS to RDF/OWL
- Tense, modality and negation
- Compositional semantics, taxonomy induction and quality representation
- Periphrastic relations
From DRS to RDF/OWL

- The core of FRED takes Discourse Representation Structures as input (based on Hans Kamp’s DRT)
- DRS informally called boxes (graphical representation)
- Discourse referents and conditions (their interpretation)
- Subset of FOL, arbitrary entities, Neo-Davidsonian events
- DRSs can become very complex as natural language expression are rarely simple
- Boxer

People love movies
1. Boxes only have a syntactic role in Boxer’s result, meaning that FRED only needs to focus on representing their content and linking them to the rest.

2. Boxes provide a unified relation to a complex state of affair (usually expressed in the text by a copula), meaning that they have their own semantics to be represented.
People love movies
Valentina is a researcher

Valentina is happy

Box pattern #2
Valentina is Gianni’s fifth daughter, from his second marriage
Complex boxes

- Boxes can be composed based on *and*, *or*, *entails*
- They can be nested, negated, contain several events or none
- Vocabularies:
  - Boxer taxonomy of types (boxer:)
  - Situations, formal relations and other properties (boxing:)
Tense, modality and negation

Tense: Time intervals + Allen’s algebra
Modality and negation: lightweight RDF semantics

Rahm Emanuel says he won’t resign over police shooting
Lightweight negation, why?

- Ambiguous negation scope, same syntactic structure

John did not go to school by car

\neg \exists e \text{go}(e, John, s, c) \land \text{Event}(e) \land \text{School}(s) \land \neg \text{Car}(c)

\exists e \text{go}(e, John, s, c) \land \text{Event}(e) \land \text{School}(s) \land \neg \text{Car}(c)
Rahm Emanuel says he won’t resign over police shooting
Modality: lightweight representation

- No formal constructs in OWL for modality
- boxing:Modality
  - boxing:Necessary (will, must, should, etc.)
  - boxing:Possible (may, might, etc.)

:resign_1 boxing:hasTruthValue boxing:False ;

  boxing:hasModality boxing:Necessary .

“will not resign”
Rahm Emanuel says he won’t resign over police shooting
Compositional semantics, taxonomy induction, and quality representation

- Compound terms: FRED recognises two main box patterns
  - *nn* dependency relations
  - co-referent predicates
- FRED creates a new class e.g. PoliceShooting, and uses punning for implicit associations
- The modifier ”police” of the main concept (“shooting) provides a distinguished feature to the more specific class *(differentia specifica)*

*Rahm Emanuel says he won’t resign over police shooting*
Rahm Emanuel says he won’t resign over police shooting
Compositional semantics: adjectives as differentia specifica

- Adjectives have an unpredictable and unstable semantics (Morzycki’s sectivity)
- Four main representation patterns implemented based on an extension of Morzycky’s theory

Compositional semantics: adjectives as differentia specifica

- **Canadian surgeon (intersective)**
  
  :surgery_1 a :CanadianSurgeon .
  :CanadianSurgeon rdfs:subClassOf :Surgeon .
  :surgery_1 dul:hasQuality :Canadian .
  :CanadianSurgeon hasQuality :Canadian .

- **Skilful surgeon (subsective)**
  
  :surgery_1 a :SkilfulSurgeon .
  :SkilfulSurgeon rdfs:subClassOf :Surgeon .
  :SkilfulSurgeon hasQuality :Skilful .

- **Alleged surgeon (modal)**
  
  :surgery_1 a :AllegedSurgeon .
  :AllegedSurgeon hasModality :Alleged .

- **Fake surgeon (privative)**
  
  :AllegedSurgeon hasQuality :Fake .
  (:AllegedSurgeon owl:disjointWith :Surgeon .) optional inference
Periphrastic relations

- Relations expressed with prepositions (of, with, for, in, etc.)
- Naming object properties with prepositions only is typically ambiguous
  - sister of, survivor of
- FRED identifies the noun to be associated with the preposition, and generates a meaningful label

```plaintext
fred:survivor_1
  fred:survivorOf fred:expedition_1 ;
  rdf:type fred:Survivor .
fred:expedition_1 rdf:type fred:Expedition .
```
Linking

- Co-reference resolution and role propagation
  - Used for merging nodes
- Named Entity Recognition
  - Used for identifying elements that should have a corresponding named entity in the knowledge graph
- Entity linking
  - owl:sameAs axioms (DBpedia)
  - Graph enrichment with types from DBpedia, schema.org, DOLCE-Zero (via OntoWordNet), etc.
- Word sense disambiguation
  - owl:equivelentClass, rdfs:subClassOf, owl:sameAs
  - WordNet, Babelnet, Wordnet Supersenses, Framester

```
curl https://lipn.univ-paris13.fr/framester/en/wfd_json/sentence -d "data=Tigers once ranged widely across Asia, from Turkey in the west to the eastern coast of Russia.:t" -X PUT
```
Other capabilities

- **Multilingualism**
  - 48 languages (translates through English)
  - Bing Translation APIs (see also Sheldon tool)

- **Textual annotation grounding**
  - Annotations of text fragment to corresponding graph elements
  - Earmark and NIF used as vocabularies
Outline

• Overview of FRED

• Main mappings and heuristics (the method behind)

• FRED pipeline (implementation)

• Quality, importance, impact (evaluation)

• Conclusion
Outline

• Overview of FRED
• Main mappings and heuristics (the method behind)
• FRED pipeline (implementation)
• Quality, importance, impact (evaluation)
• Conclusion
Quality

- Ability to perform specific knowledge extraction tasks
- Performance as a semantic middleware, by evaluating applications built on top of it
- Motif-based evaluation
## Knowledge extraction tasks: tool comparison

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

### OKE motifs: evaluation results

<table>
<thead>
<tr>
<th>Motif</th>
<th>Sample distribution</th>
<th>As FRED RDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>BINARY</td>
<td>4.72</td>
<td>A binary relation constructed (projected) out of an n-ary one</td>
</tr>
<tr>
<td>EQUIVALENCE</td>
<td>13.65</td>
<td>owl:equivalentClass</td>
</tr>
<tr>
<td>IDENTITY</td>
<td>4.29</td>
<td>owl:inverseAn</td>
</tr>
<tr>
<td>MODALITY</td>
<td>0.6</td>
<td>boxing:hasModality</td>
</tr>
<tr>
<td>N-ARY</td>
<td>19.31</td>
<td>Any subgraph (degree=*) having a different or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>boxing:hasSituation as root</td>
</tr>
<tr>
<td>NEGATION</td>
<td>0.8</td>
<td>boxing:hasTruthValue, applied to n-ary</td>
</tr>
<tr>
<td>PROPERTY</td>
<td>23.43</td>
<td>structures</td>
</tr>
<tr>
<td>SUBCLASS</td>
<td>21.46</td>
<td>Any property assertion</td>
</tr>
<tr>
<td>TYPE</td>
<td>11.85</td>
<td>rdf:type</td>
</tr>
<tr>
<td>ALL</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>W on articles</td>
<td>0.86 p=&lt;0.0001</td>
</tr>
<tr>
<td>W on definitions</td>
<td>0.74 p=&lt;0.0001</td>
</tr>
<tr>
<td>W on news</td>
<td>0.62 p=&lt;0.0001</td>
</tr>
<tr>
<td>W on tweets</td>
<td>0.65 p=&lt;0.0001</td>
</tr>
<tr>
<td>W globally</td>
<td>0.73 p=&lt;0.0001</td>
</tr>
<tr>
<td>Avg. Pearson Corr.</td>
<td>0.73</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class of triples</th>
<th>Precision r1</th>
<th>Recall r1</th>
<th>F1 r1</th>
<th>Precision r2</th>
<th>Recall r2</th>
<th>F1 r2</th>
<th>Precision r3</th>
<th>Recall r3</th>
<th>F1 r3</th>
<th>Avg. Precision</th>
<th>Avg. Recall</th>
<th>Avg. F1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Global without structure triples</strong></td>
<td>0.84</td>
<td>0.92</td>
<td>0.88</td>
<td>0.87</td>
<td>0.92</td>
<td>0.89</td>
<td>0.88</td>
<td>0.92</td>
<td>0.9</td>
<td>0.86</td>
<td>0.92</td>
<td>0.89</td>
</tr>
<tr>
<td>BINARY</td>
<td>0.76</td>
<td>0.95</td>
<td>0.85</td>
<td>0.91</td>
<td>0.98</td>
<td>0.94</td>
<td>0.89</td>
<td>0.98</td>
<td>0.93</td>
<td>0.85</td>
<td>0.97</td>
<td>0.91</td>
</tr>
<tr>
<td>EQUIVALENCE</td>
<td>0.64</td>
<td>0.77</td>
<td>0.7</td>
<td>0.74</td>
<td>0.68</td>
<td>0.71</td>
<td>0.74</td>
<td>0.68</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>IDENTITY</td>
<td>0.9</td>
<td>0.87</td>
<td>0.88</td>
<td>0.95</td>
<td>0.98</td>
<td>0.96</td>
<td>0.92</td>
<td>0.98</td>
<td>0.95</td>
<td>0.92</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>MODALITY</td>
<td>0.88</td>
<td>1</td>
<td>0.93</td>
<td>0.79</td>
<td>1</td>
<td>0.88</td>
<td>0.88</td>
<td>1</td>
<td>0.93</td>
<td>0.85</td>
<td>1.00</td>
<td>0.91</td>
</tr>
<tr>
<td>N-ARY</td>
<td>0.88</td>
<td>0.89</td>
<td>0.88</td>
<td>0.89</td>
<td>0.93</td>
<td>0.91</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.90</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td>NEGATION</td>
<td>0.86</td>
<td>1</td>
<td>0.92</td>
<td>0.86</td>
<td>1</td>
<td>0.92</td>
<td>0.86</td>
<td>1</td>
<td>0.92</td>
<td>0.86</td>
<td>1.00</td>
<td>0.92</td>
</tr>
<tr>
<td>PROPERTY</td>
<td>0.85</td>
<td>0.94</td>
<td>0.89</td>
<td>0.84</td>
<td>0.95</td>
<td>0.89</td>
<td>0.85</td>
<td>0.95</td>
<td>0.9</td>
<td>0.85</td>
<td>0.95</td>
<td>0.89</td>
</tr>
<tr>
<td>SUBCLASS</td>
<td>0.88</td>
<td>1</td>
<td>0.94</td>
<td>0.9</td>
<td>1</td>
<td>0.96</td>
<td>0.93</td>
<td>1</td>
<td>0.96</td>
<td>0.91</td>
<td>1.00</td>
<td>0.95</td>
</tr>
<tr>
<td>TYPE</td>
<td>0.9</td>
<td>0.97</td>
<td>0.93</td>
<td>0.93</td>
<td>0.96</td>
<td>0.94</td>
<td>0.92</td>
<td>0.96</td>
<td>0.94</td>
<td>0.92</td>
<td>0.96</td>
<td>0.94</td>
</tr>
<tr>
<td>ARTICLES</td>
<td>0.85</td>
<td>0.9</td>
<td>0.87</td>
<td>0.85</td>
<td>0.91</td>
<td>0.88</td>
<td>0.86</td>
<td>0.91</td>
<td>0.88</td>
<td>0.85</td>
<td>0.91</td>
<td>0.88</td>
</tr>
<tr>
<td>DEFINITIONS</td>
<td>0.81</td>
<td>0.91</td>
<td>0.86</td>
<td>0.88</td>
<td>0.93</td>
<td>0.89</td>
<td>0.89</td>
<td>0.94</td>
<td>0.91</td>
<td>0.86</td>
<td>0.93</td>
<td>0.89</td>
</tr>
<tr>
<td>NEWS</td>
<td>0.87</td>
<td>0.97</td>
<td>0.92</td>
<td>0.92</td>
<td>0.93</td>
<td>0.92</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.91</td>
<td>0.94</td>
<td>0.92</td>
</tr>
<tr>
<td>TWEETS</td>
<td>0.79</td>
<td>0.89</td>
<td>0.84</td>
<td>0.83</td>
<td>0.92</td>
<td>0.87</td>
<td>0.85</td>
<td>0.92</td>
<td>0.88</td>
<td>0.82</td>
<td>0.91</td>
<td>0.86</td>
</tr>
<tr>
<td>EQUIVALENCE with WSD only</td>
<td>0.57</td>
<td>0.44</td>
<td>0.5</td>
<td>0.7</td>
<td>0.43</td>
<td>0.53</td>
<td>0.7</td>
<td>0.43</td>
<td>0.53</td>
<td>0.66</td>
<td>0.43</td>
<td>0.52</td>
</tr>
<tr>
<td>EQUIVALENCE with NER only</td>
<td>0.81</td>
<td>0.22</td>
<td>0.35</td>
<td>0.83</td>
<td>0.32</td>
<td>0.32</td>
<td>0.82</td>
<td>0.2</td>
<td>0.32</td>
<td>0.82</td>
<td>0.21</td>
<td>0.33</td>
</tr>
<tr>
<td>GLOBAL: without WSD</td>
<td>0.87</td>
<td>0.91</td>
<td>0.89</td>
<td>0.89</td>
<td>0.91</td>
<td>0.9</td>
<td>0.9</td>
<td>0.92</td>
<td>0.91</td>
<td>0.89</td>
<td>0.91</td>
<td>0.9</td>
</tr>
<tr>
<td>TYPE with structure triples</td>
<td>0.96</td>
<td>0.99</td>
<td>0.97</td>
<td>0.97</td>
<td>0.99</td>
<td>0.98</td>
<td>0.97</td>
<td>0.98</td>
<td>0.98</td>
<td>0.97</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td>OVERALL (incl. Structure triples)</td>
<td>0.87</td>
<td>0.94</td>
<td>0.9</td>
<td>0.9</td>
<td>0.96</td>
<td>0.93</td>
<td>0.91</td>
<td>0.96</td>
<td>0.93</td>
<td>0.89</td>
<td>0.95</td>
<td>0.92</td>
</tr>
<tr>
<td>OVERALL, but without WSD</td>
<td>0.9</td>
<td>0.94</td>
<td>0.92</td>
<td>0.92</td>
<td>0.96</td>
<td>0.94</td>
<td>0.92</td>
<td>0.96</td>
<td>0.94</td>
<td>0.91</td>
<td>0.95</td>
<td>0.93</td>
</tr>
</tbody>
</table>

FRED as a semantic middleware

- Automatic entity typing (Tìpalo)
- Sentiment analysis (Sèntilo)
- Extraction of link semantics (Lègalo)
- Knowledge reconciliation (Mèrgilo)


Issues and Open challenges

- Partial accuracy of NLP components may lead to global errors

N/V POS tagging (major source of errors)

*David Moyes shares Manchester United fans' frustration*

Complex multiword extraction

*Myeloid hepatosplenomegaly is an enlargement of liver and kidney due to myelofibrosis*

Coordination

*Aristotle was a Greek philosopher, a student of Plato and teacher of Alexander the Great*

Plural co-reference

*When Carol helps Bob and Bob helps Carol, they can accomplish any task.*
Issues and Open challenges

- Semantics besides literal interpretation
  - The path descended abruptly
  - The road runs along the coast for two hours
  - The fence zigzags from the plateau to the valley
  - The highway crawls through the city
  - The road leads us to Bordeaux

- Need for type coercion to satisfy hidden frame
  - Sometimes an inversion of roles: the path descends because it can be descended
  - Road is an object that “can be followed as an indication” to our destination
  - Fence is an object whose shape “can be followed by zigzaging”
  - Highway is a path that “can be crawled”, therefore the crawling frame here is descriptive of a state, not of an action
Outline

• Overview of FRED
• Main mappings and heuristics (the method behind)
• FRED pipeline (implementation)
• Quality, importance, impact (evaluation)
• Conclusion
Conclusion

- The final goal of FRED is to support Natural Language Understanding (NLU)

- Currently, FRED responds to a very important challenge: to leverage existing NLP approaches to obtain a unified, formal knowledge graph including both facts and concepts, expressed by a natural language text

- Open and modular architecture, now integrated to the large Framester hub

- NLU is super complex: just scratching the surface

- Need to hybridise knowledge-based and learning techniques
Issues and Open challenges

• Semantics besides literal interpretation
  • *The path descended abruptly*
  • *The road runs along the coast for two hours*
  • *The fence zigzags from the plateau to the valley*
  • *The highway crawls through the city*
  • *The road leads us to Bordeaux*

• Need for type coercion to satisfy hidden frame
  • Sometimes an inversion of roles: the *path* descends because it can be descended
  • *Road* is an object that “can be followed as an indication” to our destination
  • *Fence* is an object whose shape “can be followed by zigzagging”
  • *Highway* is a path that “can be crawled”, therefore the crawling frame here is descriptive of a state, not of an action
Issues and Open challenges

• The recognition of cultural framing out of real world facts as in political discourse
• requires the extraction, representation, and reasoning over high-level frames (attitudes, values, metaphors)
Issues and Open challenges

- The accurate extraction of implicit discourse relations and conventional implicatures
  - requires background knowledge and reasoning on that knowledge in a way close to the appropriate discourse level
Community feedback (some)

- stackoverflow.com: 
  http://tinyurl.com/qa2dyfj,  
  http://tinyurl.com/o993scy

- answers.semanticweb.com:  
  http://tinyurl.com/n6pzpot,  
  http://tinyurl.com/kb8564w

- Wikipedia entry for Knowledge Extraction:  
  http://en.wikipedia.org/wiki/Knowledge_extraction

- A study about dealing with Big Data by UN Economic Commission for Europe  
  http://tinyurl.com/ml6ystn
In academic publication, for reference to FRED please cite:


Other relevant references related to the FRED project:


References


