From Hyperlinks to Semantic Web Properties using Open Knowledge Extraction

Valentina Presutti
Aldo Gangemi, Andrea Nuzzolese, Diego Reforgiato, Sergio Consoli
STLab-ISTC, CNR, Italy
The current Web vs the Semantic Web (Marja-Riitta Koivunen and Eric Miller 2001)

resources (web documents) and links

Resources and links that can have explicit types
Hyperlinks are **pragmatic traces** of semantic relations between two entities.

“McCarthy often commented on world affairs on the [Usenet forums](https://en.wikipedia.org/wiki/Usenet).”

Texts surrounding hyperlinks are **linguistic traces** of semantic relations between two entities.
He coined the term "artificial intelligence" (AI), developed the Lisp programming language family, significantly influenced the design of the ALGOL programming language, popularized timesharing, and was very influential in the early development of AI.
He coined the term "artificial intelligence" (AI), developed the Lisp programming language family, significantly influenced the design of the ALGOL programming language, popularized timesharing, and was very influential in the early development of AI.
He coined the term "artificial intelligence" (AI), developed the Lisp programming language family, significantly influenced the design of the ALGOL programming language, popularized timesharing, and was very influential in the early development of AI.
He coined the term "artificial intelligence" (AI), developed the Lisp programming language family, significantly influenced the design of the ALGOL programming language, popularized timesharing, and was very influential in the early development of AI.
He coined the term "artificial intelligence" (AI), developed the Lisp programming language family, significantly influenced the design of the ALGOL programming language, popularized timesharing, and was very influential in the early development of AI.
He coined the term "artificial intelligence" (AI), developed the Lisp programming language family, significantly influenced the design of the ALGOL programming language, popularized timesharing, and was very influential in the early development of AI.
John and Mary went to ISWC 2017, together
John and Mary went to ISWC 2017, together

John goes ISWC 2017
John and Mary went to ISWC 2017, together
John and Mary went to ISWC 2017, together
binary relations are good for interaction
querying: developers don’t want to struggle with n-ary relations
robots: to answer questions they need strategies for de-reifying n-ary relations
• and generate language out of them
similar intuition in more recent FrameBase ReDer rules
George E. Krug graduated from Lafayette College in Easton, Pennsylvania, in the class of 1884. He went on to study architecture in Philadelphia, at the Fine Arts Institute of the University of Pennsylvania.
George E. Krug graduated from Lafayette College in Easton, Pennsylvania, in the class of 1884. He went on to study architecture in Philadelphia, at the Fine Arts Institute of the University of Pennsylvania.
Relation extraction

- Distant supervision: maps textual sentences to linked data triples
  - no support for $n$-ary relation
  - bound to already defined relation types

  “Paris is the capital city of France”
  [dbpedia:Paris dbpedia-owl:country dbpedia:France]

  “John Stigall received a Bachelor of arts”

- Open Information Extraction: creates resources of triples composed of text fragments (subject, relational phrase, object)
  - not directly reusable as linked data

  “John Stigall received a Bachelor of arts”
  [(John Stigall, received, a Bachelor of arts)]

Open Challenges:
- discovery of new relations (both facts and relation types)
- methods for automatically formalising discovered knowledge
- usable label generation for new discovered relations (abstractive summarisation)
### OIE vs OKE

"John Stigall received a Bachelor of arts from the State University of New York at Cortland"

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Stigall</td>
<td>received</td>
<td>a Bachelor of arts</td>
<td>extractive</td>
</tr>
<tr>
<td>John Stigall</td>
<td>received</td>
<td>from the State University of New York at Cortland</td>
<td>extractive</td>
</tr>
<tr>
<td>dbpedia:John_Stigall</td>
<td>myprop:receives</td>
<td>dbpedia:Bachelor_of_Arts</td>
<td>abstractive</td>
</tr>
<tr>
<td></td>
<td>AcademicDegree</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dbpedia:John_Stigall</td>
<td>myprop:receives</td>
<td>dbpedia:State_University_of_New_York</td>
<td>abstractive</td>
</tr>
<tr>
<td></td>
<td>AcademicDegree From</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
“John Stigall received a Bachelor of arts from the State University of New York at Cortland “

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Stigall</td>
<td>received</td>
<td>a Bachelor of arts</td>
<td>extractive</td>
</tr>
<tr>
<td>John Stigall</td>
<td>received</td>
<td>from the State University of New York at Cortland</td>
<td>extractive</td>
</tr>
<tr>
<td>myprop:receive_1</td>
<td>role:Agent</td>
<td>dbpedia:John_Stigall</td>
<td>abstractive</td>
</tr>
<tr>
<td>myprop:receive_1</td>
<td>role:Theme</td>
<td>dbpedia:Bachelor_of_Arts</td>
<td>abstractive</td>
</tr>
<tr>
<td>myprop:receive_1</td>
<td>role:Source</td>
<td>dbpedia:Univ_NYC_Cortland</td>
<td>abstractive</td>
</tr>
</tbody>
</table>
Let’s see how it works in detail...
“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.”
Relevant Relation Assessment

"Joey Foster Ellis has published on The New York Times, and The Wall Street Journal."

\[ \varphi_s(e_{subj}, e_{obj}) \]

\( \varphi_s(\text{Joey Foster Ellis, The New York Times}) \) \( \checkmark \)

\( \varphi_s(\text{The Wall Street Journal, The New York Times}) \) \( \times \)
"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."
“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.”

Necessary condition: $\varphi_s(e_{subj}, e_{obj}) \Rightarrow \exists P(v_{subj}, v_{obj})$
“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.”
Relevant relation assessment

“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.”

Let $f$ be a node of $G$ such that $\text{dul:} \text{Event}(f)$

Role $\equiv \{\rho_1, \ldots, \rho_n\}$, the set of possible roles participating in $f$

AgRole $\equiv \{\rho_{m_1}, \ldots, \rho_{m_m}\}$, the set of VerbNet agentive roles

AgRole $\subseteq$ Role

Sufficient condition:

$\varphi_s(e_{subj}, e_{obj}) \iff$

$\exists P (v_{subj}, v_{obj})$ and $\exists f \in Pset_{subj, obj}$ such that $\text{dul:} \text{Event}(f)$ and $\rho(f, v_{subj}) \in \text{AgRole}$
“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.”
“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.”
“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.”
“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.”
“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.”
"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."

Semantic Web triples and properties generation
Usable predicate generation

\[ \varphi_s(Joey Foster Ellis, The New York Times) \]

\[ \Lambda \equiv \{\lambda_1, ..., \lambda_n\} \]

labels for \( \varphi_s(e_{\text{subj}}, e_{\text{obj}}) \)

defined for a LOD vocabulary

(\text{has}) published on
(\text{has}) published on journal
(\text{is}) author for
writes for

\[ \lambda \sim \lambda_i, \lambda_i \in \Lambda \]

publish on
“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.”
"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."

Semantic Web triples and properties generation

- Subject: Rico Lebrun
- Object: visual arts, Chouinard Art Institute, Disney Studios, Michelangelo, Goya, Picasso
- Predicate: taught, influenced, affinity
Seminar Web triples and properties generation

“**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**.”
“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.”

Semantic Web triples and properties generation

Subject

Object

teach  art  at

Subject

RNA:Actor1 -> “with”
RNA:Actor2 -> “with”
RNA:Beneficiary -> “for”
RNA:Instrument -> “with”
RNA:Destination -> “to”
RNA:Topic -> “about”
RNA:Source -> “from”
“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.”
“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.”
“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.”
“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.”

$\phi_s(Joey\ Foster\ Ellis,\ The\ New\ York\ Times)$

dbpedia:Joey_Foster_Ellis

legalo:publishOn a owl:ObjectProperty;
   rdfs:range ... ;
   rdfs:domain ... ;
... .
“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.”

```
s:teachAbout a owl:ObjectProperty ;
  rdfs:subPropertyOf fred:Teach;
  rdfs:domain wibi:Artist ;
  rdfs:range wibi:Art ;

grounding:definedFromFormalRepresentation
  fred-graph:a6705cedbf9b53d10bbcdedaa3be9791da0a9e94 ;
grounding:derivedFromLinguisticEvidence s:linguisticEvidence ;
owl:propertyChainAxiom([ owl:inverseOf s:AgentTeach ] s:TopicTeach) .
```

alignment to existing LOD vocabularies
Legalo: Evaluation sample

\[C_{\text{rel-extraction}}^*: 5 \text{ datasets each dedicated to a specific relation}\]

```json
{"pred": "/people/person/education./education/education/institution", "sub": "/m/0g9zjv3", "obj": "/m/0ckrjh", "evidences": ["url":"http://en.wikipedia.org/wiki/Dmitry_Chernyakov", "snippet": "Dmitry was born in Moscow. In 1993 he graduated from Russian Academy of Theatre Arts as stage director. He started his career in the Russian Drama Theatre of Lithuania in Vilnius. Then he directed opera and drama in many major Russian cities: Moscow, Saint Petersburg, Novosibirsk, Omsk, Samara, Kazan and others. He usually creates scenic design and stage clothes for his plays."], "judgments": [{"rater":"1264223381988340244","judgment":"yes"}, {"rater":"6968726908160095830","judgment":"yes"}, {"rater":"6063741883945424276","judgment":"yes"}, {"rater":"5346153820624061638","judgment":"yes"}, {"rater":"12022408018620867151","judgment":"yes"}]
```

*https://code.google.com/p/relation-extraction-corpus/downloads/list corpus for relation extraction*
Legalo: Evaluation sample

$C_{\text{institution}}$: random sample of 130 snippets (one entity pair) from $C_{\text{rel-extraction}}$ for “attending or graduating from an institution”
Legalo produced always an output, either $p$ or “no relation”

$C_{\text{education}}$: random sample of 130 snippets (one entity pair) from $C_{\text{rel-extraction}}$ for “obtaining a degree of education”
Legalo produced always an output, either $p$ or “no relation”

$C_{\text{general}}$: random sample of 60 snippets from all $C_{\text{rel-extraction}}$ datasets then broken into 186 single sentences with at least one entity pair.
Legalo produced 867 results, 262 $p$ and 605 “no relation”
<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Task</th>
<th>#workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis 1</td>
<td>Task 1,2</td>
<td>35</td>
</tr>
<tr>
<td>Hypothesis 2</td>
<td>Task 3 (institution)</td>
<td>10</td>
</tr>
<tr>
<td>Hypothesis 2</td>
<td>Task 3 (education)</td>
<td>18</td>
</tr>
<tr>
<td>Hypothesis 2</td>
<td>Task 4</td>
<td>19</td>
</tr>
<tr>
<td>Hypothesis 2</td>
<td>Task 5</td>
<td>12</td>
</tr>
</tbody>
</table>
Hypothesis 1: Relevant relation assessment

Legalo is able to assess if, given a sentence $s$, a relevant relation exists which holds between two entities, according to the content of $s$:

$$\exists \varphi. \varphi_s(e_{subj}, e_{obj})$$
Crowdsourcing tasks for Relevant Relation Assessment

Task 1:
Assessing if a sentence $s$ is an evidence of a referenced relation (i.e. either “institution” or “education”) between two entities, mentioned in $s$.

Based on data from $C_{\text{institution}}$ and $C_{\text{education}}$, respectively

Task 2:
Assessing if a sentence $s$ is an evidence of any relation between two given entities mentioned in $s$.

Based on data from $C_{\text{general}}$

At least 3 raters with $t>0.70$
Answer could be “yes” or “no”
### Evaluation results for Relevant Relation Assessment

The confidence value expresses a weighted value for inter-rater agreement, by rater trust scores.

<table>
<thead>
<tr>
<th>Task</th>
<th>Relation</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
<th>Accuracy</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Any</td>
<td>0.83</td>
<td>0.92</td>
<td>0.87</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>1</td>
<td>Education</td>
<td>0.95</td>
<td>0.91</td>
<td>0.93</td>
<td>0.87</td>
<td>0.96</td>
</tr>
<tr>
<td>1</td>
<td>Institution</td>
<td>0.93</td>
<td>0.90</td>
<td>0.91</td>
<td>0.84</td>
<td>0.94</td>
</tr>
</tbody>
</table>
Hypothesis 2: Usable predicate generation

Legalo is able to generate a usable predicate \( \lambda \) for a relevant relation \( \varphi_s \) between two entities, expressed in a sentence \( s \): given \( \lambda \), a label generated by Legalo for \( \varphi_s \), and \( \lambda_i \), a label generated by a human for \( \varphi_s \), the following holds:

\[ \lambda \sim \lambda_i, \lambda_i \in \Lambda \]

\[ \Lambda \equiv \{ \lambda_1, \ldots, \lambda_n \} \]

labels for \( \varphi_s(e_{\text{subj}}, e_{\text{obj}})_i \) defined for a LOD vocabulary
Crowdsourcing tasks for Usable predicate generation

Task 3:

Judging if λ generated by a machine expresses is a good summarisation of a specific relation (i.e. either “institution” or “education”) between two given entities mentioned in s, according to the evidence provided by s.

Based on data from $C_{\text{institution}}$ and $C_{\text{education}}$, respectively

Task 4:

judging if a predicate λ is a good summarisation of a relation expressed by the content of s, between two given entities mentioned in s, according to the evidence provided by s.

Based on data from $C_{\text{general}}$

At least 3 raters with t>0.70

Answer could be Agree, Partly Agree, Disagree
Evaluation results for Usable Predicate Generation

<table>
<thead>
<tr>
<th>Task</th>
<th>Relation</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
<th>Accuracy</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Education</td>
<td>0.92</td>
<td>0.91</td>
<td>0.91</td>
<td>0.85</td>
<td>0.80</td>
</tr>
<tr>
<td>3</td>
<td>Institution</td>
<td>0.65</td>
<td>0.91</td>
<td>0.76</td>
<td>0.62</td>
<td>0.59</td>
</tr>
<tr>
<td>3 (high confidence only)</td>
<td>Institution</td>
<td>0.74</td>
<td>0.89</td>
<td>0.81</td>
<td>0.68</td>
<td>0.71</td>
</tr>
<tr>
<td>4</td>
<td>Any</td>
<td>0.68</td>
<td>0.90</td>
<td>0.78</td>
<td>0.71</td>
<td>0.64</td>
</tr>
<tr>
<td>4 (high confidence only)</td>
<td>Any</td>
<td>0.73</td>
<td>0.87</td>
<td>0.80</td>
<td>0.75</td>
<td>0.76</td>
</tr>
</tbody>
</table>

The confidence value expresses a weighted value for inter-rater agreement, by rater trust scores.

Agree = 1, Partly Agree = 0.5, Disagree = 0
Crowdsourcing tasks for Usable predicate generation

Task 5:

Creating a phrase $\lambda$ that summarises the relation expressed by the content of $s$, between two given entities mentioned in $s$, according to the evidence provided by $s$.

Based on data from $C_{\text{general}}$

At least 3 raters with $t>0.60$  
Answer was open
Evaluation results for Usable Predicate Generation

- Similarity score between human created \( \{\lambda_i\} \) and Legalo \( \lambda_i \), for a \( \varphi_s(e_{subj}, e_{obj}) \)
- Jaccard distance measure (string similarity)
- Semantic similarity measure based on the SimLibrary framework*

<table>
<thead>
<tr>
<th>Task</th>
<th>Relation</th>
<th>Jaccard [0,1]</th>
<th>SimLibrary [0,1]</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Any</td>
<td>0.63</td>
<td>0.80</td>
<td>0.59</td>
</tr>
</tbody>
</table>

* G. Pirrò and J. Euzenat. A feature and information theoretic framework for semantic similarity and relatedness. ISWC2010
Conclusion and next steps

• a method to extract the meaning of hyperlinks
• and express it as de-reification of n-ary relations
• ability to identify relevant binary relations
• ability to generate usable labels
• useful for interaction tasks
  • e.g. querying, language generation

Next

• reconciliation strategies
• robot’s question answering
Thanks for your attention

Questions?

http://wit.istc.cnr.it/stlab-tools/legalo
References


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

http://wit.istc.cnr.it/stlab-tools/fred
"Joey Foster Ellis has published on The New York Times, and The Wall Street Journal."

Set of entity pairs

\[ \{(e_{subj}, e_{obj})_i\} \]

Relevant relations evidenced by \( s \)

\[ \varphi_s(e_{subj}, e_{obj})_i \]

\[ \varphi_s(\text{Joey Foster Ellis, The New York Times}) \]
Method and implementation
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.”
“The Black Hand might not have decided to barbarously assassinate Franz Ferdinand after he arrived in Sarajevo on June 28th, 1914”
Open Knowledge Extraction

- **unsupervised** -> no need of huge annotated corpora for training
- **open-domain** -> not bound to specific domains
- **abstractive** -> models the text, uses external knowledge resources, uses summarisation and language generation techniques
- **producing formalised output** -> linked data knowledge graph, (de-)reification of expressed relations
### OIE vs OKE

“John Stigall received a Bachelor of arts from the State University of New York at Cortland“

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Stigall</td>
<td>received</td>
<td>a Bachelor of arts</td>
<td>extractive</td>
</tr>
<tr>
<td>John Stigall</td>
<td>received</td>
<td>from the State University of New York at Cortland</td>
<td>extractive</td>
</tr>
<tr>
<td>dbpedia:John_Stigall</td>
<td>myprop:receives AcademicDegree</td>
<td>dbpedia:Bachelor_of_Arts</td>
<td>abstractive</td>
</tr>
<tr>
<td>dbpedia:John_Stigall</td>
<td>myprop:receives AcademicDegree From</td>
<td>dbpedia:State_University_of_New_York</td>
<td>abstractive</td>
</tr>
<tr>
<td>Subject</td>
<td>Predicate</td>
<td>Object</td>
<td>Approach</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------</td>
<td>------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>John Stigall</td>
<td>received</td>
<td>a Bachelor of arts</td>
<td>extractive</td>
</tr>
<tr>
<td>John Stigall</td>
<td>received</td>
<td>from the State University of New York at Cortland</td>
<td>extractive</td>
</tr>
<tr>
<td>myprop:receive_1</td>
<td>role:Agent</td>
<td>dbpedia:John_Stigall</td>
<td>abstractive</td>
</tr>
<tr>
<td>myprop:receive_1</td>
<td>role:Theme</td>
<td>dbpedia: Bachelor_of_Arts</td>
<td>abstractive</td>
</tr>
<tr>
<td>myprop:receive_1</td>
<td>role:Source</td>
<td>dbpedia: Univ_NYC_Cortland</td>
<td>abstractive</td>
</tr>
</tbody>
</table>
OKE tools

Performing unsupervised, open domain, and abstractive knowledge extraction from text for producing directly usable machine readable data

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

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

http://wit.istc.cnr.it/stlab-tools/legalo/wikipedia
Method and implementation
Method and implementation
Method and implementation

Frame-based formal representation of s

Relation assessment

Any Natural Language sentence

FRED machine reader

FRED graph

Entity pair selection

Relation Assessment (*)

Hyperlink/Entity Identification(*)

Formaliser (*)

Semantic Web property generator (*)

Property matcher (*)

RDF/OWL Writer

RDF/OWL triples (*)

Watson

LOV

NELL
Method and implementation

Label generation (extraction+abstraction)

Frame-based formal representation of \( s \)

Relation assessment

Any Natural Language sentence

FRED machine reader

FRED graph

Entity pair selection

Relation Assessment (*)

Hyperlink/Entity Identification(*)

Formaliser (*)

Semantic Web property generator (*)

Property matcher (*)

RDF/OWL Writer

RDF/OWL triples (*)

Watson

LOV

NELL
Method and implementation

Frame-based formal representation of s

Label generation (extraction+abstraction)

Formalisation

Relation assessment

Hyperlink/Entity Identification (*)

Relation Assessment (*)

Entity pair selection

Any Natural Language sentence

FRED machine reader

FRED graph

Formaliser (*)

Semantic Web property generator (*)

Property matcher (*)

RDF/OWL Writer

RDF/OWL triples (*)

NELL

LOV

Watson
Method and implementation

Label generation (extraction+abstraction)

Formalisation

Alignment to existing LOD vocabularies

Relation assessment

Frame-based formal representation of $s$

Any Natural Language sentence

FRED machine reader

FRED graph

Entity pair selection

Relation Assessment (*)

Hyperlink/Entity Identification (*)

Formaliser (*)

Semantic Web property generator (*)

Property matcher (*)

RDF/OWL triples (*)

RDF/OWL Writer

Watson

LOV

NELL
Method and implementation

- **Frame-based formal representation of s**
  - Any Natural Language sentence
  - FRED machine reader
  - FRED graph

- **Label generation (extraction+abstraction)**
  - Formalisation
  - Formaliser (*)
  - Property matcher (*)

- **Alignment to existing LOD vocabularies**
  - Semantic Web property generator (*)
  - RDF/OWL Writer
  - RDF/OWL triples (*)

- **Subjects and objects are given by DBpedia pagelinks**

- **Linguistic traces of Wikipedia pagelinks**
  - Relation Assessment (*)
  - Entity pair selection
  - Hyperlink/Entity Identification(*)
“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.”
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.”
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.”
“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.”
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.”
“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.”
“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.”
“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.”

$G = (V,E), V = \{v_0, \ldots, v_n\}, E = \{edge_1, \ldots, edge_n\}$
"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."

\[ G = (V,E), \ V = \{v_0, \ldots, v_n\}, \ E = \{edge_1, \ldots, edge_n\} \]

\[ G' = (V,E') \text{ is the undirected version of } G = (V,E) \]
Binary relation assessment

“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.”

G=(V,E), V = \{v_0, ..., v_n\}, E = \{edge_1, ..., edge_n\}

G’=(V,E’) is the undirected version of G=(V,E)

v_i \in V is the node in G representing the entity e_i mentioned in s.
Binary relation assessment

“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**.”

\[ G=(V,E), \ V=\{v_0, \ldots, v_n\}, \ E=\{\text{edge}_1, \ldots, \text{edge}_n\} \]

\[ G’=(V,E’) \text{ is the undirected version of } G=(V,E) \]

\( v_i \in V \) is the node in \( G \) representing the entity \( e_i \) mentioned in \( s \).

\[ P(v_{\text{subj}},v_{\text{obj}})=[v_0,\text{edge}_1,\ldots,\text{edge}_n,v_n], \ P_{\text{set subj, obj}}=\{\text{edge}_1, v_1, \ldots, v_{n-1}, \text{edge}_n\} \]
Evaluation
Legalo-Wikipedia (EKAW 2014)

<table>
<thead>
<tr>
<th>Number of $p_{new}$</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
<th>Kendall’s W</th>
</tr>
</thead>
<tbody>
<tr>
<td>629</td>
<td>0.72</td>
<td>0.97</td>
<td>0.83</td>
<td>0.73</td>
</tr>
</tbody>
</table>

3 raters, computer science, non-familiar with Legalo, familiar with Linked Data