Improving NLP Entity Annotations via Ontological Knowledge
Context: Knowledge Extraction

Kia has hired Peter Schreyer as chief design officer.
Kia has hired Peter Schreyer as chief design officer.
Context: Knowledge Extraction

Kia has hired Peter Schreyer as chief design officer.

NLP Tasks:
- Named Entity Recognition and Classification (NERC)
Kia has hired Peter Schreyer as chief design officer.

NLP Tasks:
- Named Entity Recognition and Classification (NERC)
- Entity Linking (EL)
Context: Knowledge Extraction

Kia has hired Peter Schreyer as chief design officer.

NLP Tasks:
- Named Entity Recognition and Classification (NERC)
- Entity Linking (EL)
- Semantic Role Labeling (SRL)
...
Motivating Examples

Lincoln is based in Michigan.
Motivating Examples

Lincoln is based in Michigan.

Stanford CoreNLP
http://nlp.stanford.edu:8080/corenlp
Motivating Examples

Lincoln is based in Michigan.
Motivating Examples

Lincoln is based in Michigan.

San Jose is one of the strongest hockey team.
Motivating Examples

Lincoln is based in Michigan.

San Jose is one of the strongest hockey team.
Motivating Examples

Lincoln is based in Michigan.

San Jose is one of the strongest hockey team.
Abstracting

… token₁ token₂ token₃ token₄ token₅ token₆ ….
Abstracting
Abstracting

... token_1 token_2 token_3 token_4 token_5 token_6 ....

Task_1

Task_2

Task_n
Abstracting
Improving NLP Entity Annotations via Ontological Knowledge

Marco Rospocher

Abstracting

... token_1 token_2 token_3 token_4 token_5 token_6 ....
Abstracting

\[ a_{1,1}, a_{1,2}, \ldots, a_{1,k}, \ldots, a_{2,1}, a_{2,2}, \ldots, a_{2,i}, \ldots, a_{n,1}, a_{n,2}, \ldots, a_{n,j} \]

\[ \text{... token}_1 \text{ token}_2 \text{ token}_3 \text{ token}_4 \text{ token}_5 \text{ token}_6 \ldots \]
RESEARCH PROBLEM

How can we assess and improve the coherence of the various NLP annotations on an entity mention?
In a nutshell

ontological background knowledge

... token₁ token₂ token₃ token₄ token₅ token₆ ....

Task₁

... a₁,1 a₁,2... a₁,k

Task₂

... a₂,1 a₂,2... a₂,i

Taskₙ

... aₙ,1 aₙ,2... aₙ,j
In a nutshell

ontological background knowledge

... token₁ token₂ token₃ token₄ token₅ token₆ ....

Task₁

Task₂

Taskₙ

Marco Rospocher

Improving NLP Entity Annotations via Ontological Knowledge
Improving NLP Entity Annotations via Ontological Knowledge

In a nutshell

ontological background knowledge

Task_1

Task_2

Task_n

... token_1 token_2 token_3 token_4 token_5 token_6 ....
In a nutshell

ontological background knowledge

Task_1

Task_2

Task_n

... token_1 token_2 token_3 token_4 token_5 token_6 ....
Contributions

Marco Rospocher, Francesco Corcoglioniti
Joint Posterior Revision of NLP Annotations via Ontological Knowledge
IJCAI-18

Marco Rospocher
An Ontology-Driven Probabilistic Soft Logic Approach to Improve NLP Entity Annotations
ISWC-18
Contributions

• A concrete instantiation of the models for NERC and EL (using YAGO as ontological knowledge)

• Application of the NERC and EL models to revise the annotations of Stanford NER and DBpedia Spotlight
Contributions

Marco Rospocher, Francesco Corcoglioniti
Joint Posterior Revision of NLP Annotations via Ontological Knowledge
IJCAI-18

Marco Rospocher
An Ontology-Driven Probabilistic Soft Logic Approach to Improve NLP Entity Annotations
ISWC-18
The JPARK Model

\[ P(\alpha|m, B, K) \]
The JPARK Model

entity mention  NLP Background Knowledge

$(a_i, \ldots, a_n)$ NLP Annotations

"The" Ontological Knowledge

$P(a|m, B, K)$
The JPARC Model

The model can be described as:

$$P(a|m, B, K)$$

Where:
- $a_i, \ldots, a_n$ are NLP Annotations
- $m$ is a set of classes from $K$
- $B$ is the “The” Ontological Knowledge
- $K$ is the set of classes

The equation can be further expanded as:

$$\sum_C P(a, C|m, B, K)$$
Improving NLP Entity Annotations via Ontological Knowledge

**The JPARK Model**

\[
P(a|m, B, K) = \sum_C P(a, C|m, B, K) \cdot P(C|m, B, K) \cdot P(a|m, B, K, C)
\]

- \(P(a|m, B, K)\): Probability of an entity mention given NLP annotations, background knowledge, and ontological knowledge.
- \(P(a, C|m, B, K)\): Joint probability of an entity mention and its class.
- \(P(C|m, B, K)\): Probability of a class given NLP annotations and background knowledge.
- \(P(a|m, B, K, C)\): Probability of an entity mention given the class.

**NLP Annotations**

\((a_1, \ldots, a_n)\)

**NLP Background Knowledge**

\(K\)

**Ontological Knowledge**

\((M)\)

**Set of Classes**

\(C\)

Marco Rospocher

Improving NLP Entity Annotations via Ontological Knowledge
**The JPARK Model**

entity mention \( a_i, \ldots, a_n \) NLP Annotations

NLP Background Knowledge

“The” Ontological Knowledge

\[
P(a|m, B, K) = \sum_C P(a, C|m, B, K) \prod_i P(a_i|m, B, K, C)
\]

set of classes from \( K \) \( \{M\} \)

\( \{CP\} \)
The JPARK Model

\[
\sum_C P(a, C|m, B, K) \quad \text{\textasciitilde\textit{(CP)}}
\]

\[
P(C|m, B, K) \cdot P(a|m, B, K, C) \quad \text{\textasciitilde\textit{(CIA-1)}}
\]

\[
\prod_i P(a_i|m, B, K, C) \quad \text{\textasciitilde\textit{(CP)}}
\]

\[
\prod_i P(a_1, C|m, B, K) \quad \text{\textasciitilde\textit{(CP)}}
\]

\[
\frac{P(C|m, B, K)^n}{P(C|m, B, K)}
\]
The JPARK Model

\[ P(a|m, B, K) \]

\[ \sum_C P(a, C|m, B, K) \]

\[ \prod_i P(a_i, C|m, B, K) \]

\[ \frac{P(C|m, B, K)^{n-1}}{P(C|m, B, K)^{n-1}} \]
The JPARK Model

\[ P(C|m, B, K) \]

\[ P(a_i, C|m, B, K) \]
The JPARK Model

\[
P(C|m, B, K) \overset{(M^*)}{=} \left( \prod_i \sum_{a_i} P(a_i, C|m, B, K) \right)^{\frac{1}{n}}
\]

\[
P(a_i, C|m, B, K)
\]
\[ \Pr(C|m, B, K) \overset{(M^*)}{=} \left( \prod_i \sum_{a_i} \Pr(a_i, C|m, B, K) \right)^{\frac{1}{n}} \]

\[ \Pr(a_i, C|m, B, K) \overset{(CP)}{=} \Pr(a_i|m, B, K) \cdot \Pr(C|a_i, m, B, K) \]
The JPARK Model

\[
P(C|m, B, K) \overset{(M*)}{=} \left( \prod_i \sum_{a_i} P(a_i, C|m, B, K) \right)^{\frac{1}{n}}
\]

\[
P(a_i, C|m, B, K) \overset{(CP)}{=} P(a_i|m, B, K) \cdot P(C|a_i, m, B, K)
\]

\[
(CIA-2) \parallel
\]

\[
P(a_i|m, B)
\]
The JPARK Model

\[ P(C|m, B, K) \overset{(M^*)}{=} \left( \prod_i \sum_{a_i} P(a_i, C|m, B, K) \right)^{\frac{1}{n}} \]

\[ P(a_i, C|m, B, K) \overset{(CP)}{=} P(a_i|m, B, K) \cdot P(C|a_i, m, B, K) \]

\[ \overset{(CIA-2) \parallel}{=} \quad \overset{\parallel (CIA-3)}{=} \]

\[ P(a_i|m, B) \quad P(C|a_i, K) \]
The JPARK Model

\[ P(C|m, B, K)^{(M*)} = \left( \prod_i \sum_{a_i} P(a_i, C|m, B, K) \right)^{\frac{1}{n}} \]

\[ P(a_i, C|m, B, K)^{(CP)} = P(a_i|m, B, K) \cdot P(C|a_i, m, B, K) \]

\( (\text{CIA-2}) \parallel \)

\( (\text{CIA-3}) \parallel \)

\[ P(a_i|m, B) \]

\[ P(C|a_i, K) \]

confidence score
The JPARK Model

\[ P(C|m, B, K) \overset{(M)}{=} (\prod_i \sum_{a_i} P(a_i, C|m, B, K))^\frac{1}{n} \]

\[ P(a_i, C|m, B, K) \overset{(CP)}{=} P(a_i|m, B, K) \cdot P(C|a_i, m, B, K) \]

(CIA-2) || (CIA-3)

- \( P(a_i|m, B) \)
- \( P(C|a_i, K) \)

confidence score

learned from data
The JPARK Model

\[ P(a|m, B, K) \]

\[ P(a_i|m, B) \quad P(C|a_i, K) \]
The **JPARK** Model

\[
= \arg \max_a P(a|m, B, K)
\]

\[
P(a_i|m, B) \quad P(C|a_i, K)
\]
Ontological Background Knowledge

6,016,695 entities
Taxonomy of 568,255 classes

[Suchanek et al., 2007]
Ontological Background Knowledge

6,016,695 entities
Taxonomy of 568,255 classes

\[ \text{yago select knowledge} \quad \text{[Suchanek et al., 2007]} \quad + \quad \text{WIKIPEDIA}
\]

The Free Encyclopedia
(only ingoing links)
Estimating $P(C|a_{\text{NERC}}, K)$

Leverage a gold standard corpus $G$ annotated with NERC types and ontological classes (or EL annotations)
Estimating $P(C|a_{\text{NERC}}, K)$

Leverage a gold standard corpus $G$ annotated with NERC types and ontological classes (or EL annotations)

\[
\sim \frac{n_G(C, a_{\text{NERC}})}{\sum_{C'} n_G(C', a_{\text{NERC}})}
\]
Estimating $P(C|a_{\text{NERC}}, K)$

$$\frac{n_G(C, a_{\text{NERC}})}{\sum_{C'} n_G(C', a_{\text{NERC}})}$$
Estimating $P(C|a_{\text{NERC}}, K)$

$$\alpha \cdot P(C|K) + (1 - \alpha) \cdot P(C|a_{\text{NERC}}, G)$$

$$\| \frac{n_G(C, a_{\text{NERC}})}{\sum_{C'} n_G(C', a_{\text{NERC}})}$$
Estimating $P(C|a_{\text{NERC}}, K)$

$$
\alpha \cdot P(C|K) + (1 - \alpha) \cdot P(C|a_{\text{NERC}}, G)
$$

Prior (popularity based on entity ingoing links)
Estimating \[ P(C|a_{\text{NERC}}, K) \]

\[
\alpha \cdot P(C|K) + (1 - \alpha) \cdot P(C|a_{\text{NERC}}, G)
\]

Prior (popularity based on entity ingoing links)

Consider only class sets restricted to **popular classes**
Estimating $P(C|\alpha_{EL}, K)$

Leverage alignments between EL Knowledge Base and yago select knowledge.
Estimating $P(C|a_{EL}, K)$

Leverage alignments between EL Knowledge Base and yago*

$$1_{\{C_K(a_{EL})\}}(C) \begin{cases} 
1 & \text{entity } a_{EL} \text{ is “instance” of } C \\
0 & \text{otherwise}
\end{cases}$$

classes of the entity from linking
Application and Evaluation
Tools

- **NERC**: Stanford CoreNLP [Finkel et al., 2005]
- **EL**: DBpedia Spotlight [Daiber et al., 2013]
NERC+EL Datasets

- AIDA CoNLL-YAGO [Hoffart et al., 2011]
- MEANTIME [Minard et al., 2016]
- TAC-KBP [Ji et al., 2011]
NERC+EL Datasets

- AIDA CoNLL-YAGO [Hoffart et al., 2011]
  \[ P(C|a_{\text{NERC}}, K) \] learned from AIDA CoNLL-YAGO (train)

- MEANTIME [Minard et al., 2016]

- TAC-KBP [Ji et al., 2011]
Research Question

Does the JPARK posteriori joint revision of the annotations from Stanford NER and DBpedia Spotlight, via YAGO, improve their NERC and EL performances?
Research Question

Does the JPARK posteriori joint revision of the annotations from Stanford NER and DBpedia Spotlight, via YAGO, improve their NERC and EL performances?
## Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Type</th>
<th>Link</th>
<th>Type+Link</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P$</td>
<td>$R$</td>
<td>$F_1$</td>
</tr>
<tr>
<td>AIDA (5616)</td>
<td>standard</td>
<td>.943</td>
<td>.875</td>
</tr>
<tr>
<td></td>
<td>with JPAR</td>
<td>.950</td>
<td>.881</td>
</tr>
<tr>
<td></td>
<td>$\Delta$</td>
<td>.007</td>
<td>.006</td>
</tr>
<tr>
<td>MEANTIME (792)</td>
<td>standard</td>
<td>.882</td>
<td>.695</td>
</tr>
<tr>
<td></td>
<td>with JPAR</td>
<td>.914</td>
<td>.720</td>
</tr>
<tr>
<td></td>
<td>$\Delta$</td>
<td>.032</td>
<td>.025</td>
</tr>
<tr>
<td>TAC-KBP (4969)</td>
<td>standard</td>
<td>.911</td>
<td>.652</td>
</tr>
<tr>
<td></td>
<td>with JPAR</td>
<td>.926</td>
<td>.663</td>
</tr>
<tr>
<td></td>
<td>$\Delta$</td>
<td>.015</td>
<td>.011</td>
</tr>
</tbody>
</table>

Bold = statistical significant (approx. rand. test)
## Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Type</th>
<th>Link</th>
<th>Type+Link</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P$</td>
<td>$R$</td>
<td>$F_1$</td>
</tr>
<tr>
<td><strong>AIDA (5616)</strong></td>
<td>standard</td>
<td>.943</td>
<td>.875</td>
</tr>
<tr>
<td></td>
<td>with JPARK</td>
<td>.950</td>
<td>.881</td>
</tr>
<tr>
<td></td>
<td>$\Delta$</td>
<td><strong>.007</strong></td>
<td><strong>.006</strong></td>
</tr>
<tr>
<td><strong>MEANTIME (792)</strong></td>
<td>standard</td>
<td>.882</td>
<td>.695</td>
</tr>
<tr>
<td></td>
<td>with JPARK</td>
<td>.914</td>
<td>.720</td>
</tr>
<tr>
<td></td>
<td>$\Delta$</td>
<td><strong>.032</strong></td>
<td><strong>.025</strong></td>
</tr>
<tr>
<td><strong>TAC-KBP (4969)</strong></td>
<td>standard</td>
<td>.911</td>
<td>.652</td>
</tr>
<tr>
<td></td>
<td>with JPARK</td>
<td>.926</td>
<td>.663</td>
</tr>
<tr>
<td></td>
<td>$\Delta$</td>
<td><strong>.015</strong></td>
<td><strong>.011</strong></td>
</tr>
</tbody>
</table>

Bold = statistical significant (approx. rand. test)
### Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Type</th>
<th>Link</th>
<th>Type+Link</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P$</td>
<td>$R$</td>
<td>$F_1$</td>
</tr>
<tr>
<td><strong>AIDA (5616)</strong></td>
<td>standard</td>
<td>.943</td>
<td>.875</td>
</tr>
<tr>
<td></td>
<td>with JPARK</td>
<td>.950</td>
<td>.881</td>
</tr>
<tr>
<td></td>
<td><strong>$\Delta$</strong></td>
<td><strong>.007</strong></td>
<td><strong>.006</strong></td>
</tr>
<tr>
<td><strong>MEANTIME (792)</strong></td>
<td>standard</td>
<td>.882</td>
<td>.695</td>
</tr>
<tr>
<td></td>
<td>with JPARK</td>
<td>.914</td>
<td>.720</td>
</tr>
<tr>
<td></td>
<td><strong>$\Delta$</strong></td>
<td><strong>.032</strong></td>
<td><strong>.025</strong></td>
</tr>
<tr>
<td><strong>TAC-KBP (4969)</strong></td>
<td>standard</td>
<td>.911</td>
<td>.652</td>
</tr>
<tr>
<td></td>
<td>with JPARK</td>
<td>.926</td>
<td>.663</td>
</tr>
<tr>
<td></td>
<td><strong>$\Delta$</strong></td>
<td><strong>.015</strong></td>
<td><strong>.011</strong></td>
</tr>
</tbody>
</table>

Bold = statistical significant (approx. rand. test)
Research Question

Does the JPARK posteriori joint revision of the annotations from Stanford NER and DBpedia Spotlight, via YAGO, improve their NERC and EL performances?
Research Question

Does the JPARK posteriori joint revision of the annotations from Stanford NER and DBpedia Spotlight, via YAGO, improve their NERC and EL performances?
Contributions

- Joint Posterior Revision of NLP Annotations via Ontological Knowledge
  - Marco Rospocher, Francesco Corcoglioniti
  - IJCAI-18

- An Ontology-Driven Probabilistic Soft Logic Approach to Improve NLP Entity Annotations
  - Marco Rospocher
  - ISWC-18
Improving NLP Entity Annotations via Ontological Knowledge

Marco Rospocher
Improve NLP Entity Annotations via Ontological Knowledge

Don't miss: 21 Nov 2018
Lise Getoor Keynote!
in a nutshell (1/3)

1.2: \( \text{WorksFor}(b, c) \land \text{BossOf}(b, e) \rightarrow \text{WorksFor}(e, c) \)
in a nutshell (1/3)

1.2: \( \text{WorksFor}(b, c) \land \text{BossOf}(b, e) \rightarrow \text{WorksFor}(e, c) \)

\[ \text{weight} \]
1.2: \( \text{WorksFor}(b, c) \land \text{BossOf}(b, e) \rightarrow \text{WorksFor}(e, c) \)

**weight** variable
in a nutshell (1/3)

1.2: \( \text{WorksFor}(b, c) \land \text{BossOf}(b, e) \rightarrow \text{WorksFor}(e, c) \)

weight variable predicate
in a nutshell (1/3)

1.2: \( \text{WorksFor}(b, c) \land \text{BossOf}(b, e) \rightarrow \text{WorksFor}(e, c) \)
in a nutshell (1/3)

\[ \text{body} \]

1.2: \( \text{WorksFor}(b, c) \land \text{BossOf}(b, e) \rightarrow \text{WorksFor}(e, c) \)

weight variable predicate atom
in a nutshell (1/3)

\[
\text{body} \\
1.2: \quad \text{WorksFor}(b, c) \land \text{BossOf}(b, e) \rightarrow \text{WorksFor}(e, c) \\
\text{head}
\]

weight \quad \text{variable} \quad \text{predicate} \quad \text{atom}
in a nutshell (1/3)

1.2: \( \text{WorksFor}(b, c) \land \text{BossOf}(b, e) \rightarrow \text{WorksFor}(e, c) \)

grounding

\( \text{WorksFor}(John, FBK) \)
in a nutshell (1/3)

1.2: $\text{WorksFor}(b, c) \land \text{BossOf}(b, e) \rightarrow \text{WorksFor}(e, c)$

**grounding**

$\text{WorksFor}(\text{John}, \text{FBK})$

**soft-truth value $\in [0,1]$**
in a nutshell (1/3)

1.2: \( \text{WorksFor}(b, c) \land \text{BossOf}(b, e) \rightarrow \text{WorksFor}(e, c) \)

grounding

\[ \text{WorksFor}(\text{John}, \text{FBK}) \]

soft-truth value \( \in [0, 1] \)

Interpretation \( I : \{ \text{ground atoms} \} \rightarrow [0, 1]^n \)
in a nutshell (2/3)

Lukasiewicz t-norm/co-norm

\[ I(a_1) \land I(a_2) = \max\{I(a_1) + I(a_2) - 1, 0\} \]
\[ I(a_1) \lor I(a_2) = \min\{I(a_1) + I(a_2), 1\} \]
\[ \neg I(a_1) = 1 - I(a_1) \]
Lukasiewicz t-norm/co-norm

\[ I(a_1) \land I(a_2) = \max\{I(a_1) + I(a_2) - 1, 0\} \]
\[ I(a_1) \lor I(a_2) = \min\{I(a_1) + I(a_2), 1\} \]
\[ \neg I(a_1) = 1 - I(a_1) \]

Rule is satisfied iff \( I(\text{body}) \leq I(\text{head}) \)
Lukasiewicz t-norm/co-norm

\[ I(a_1) \land I(a_2) = \max\{I(a_1) + I(a_2) - 1, 0\} \]

\[ I(a_1) \lor I(a_2) = \min\{I(a_1) + I(a_2), 1\} \]

\[ \neg I(a_1) = 1 - I(a_1) \]

rule is satisfied iff \( I(\text{body}) \leq I(\text{head}) \)

distance to satisfaction \( d(r) = \max\{0, I(\text{body}) - I(\text{head})\} \)
in a nutshell (2/3)

Lukasiewicz t-norm/co-norm

\[ I(a_1) \land I(a_2) = \max\{I(a_1) + I(a_2) - 1, 0\} \]
\[ I(a_1) \lor I(a_2) = \min\{I(a_1) + I(a_2), 1\} \]
\[ \neg I(a_1) = 1 - I(a_1) \]

rule is satisfied iff \( I(\text{body}) \leq I(\text{head}) \)

distance to satisfaction \( d(r) = \max\{0, I(\text{body}) - I(\text{head})\} \)

\[ \text{WorksFor}(John, FBK) \land \text{BossOf}(John, Jack) \rightarrow \text{WorksFor}(Jack, FBK) \]
Improving NLP Entity Annotations via Ontological Knowledge

Lukasiewicz t-norm/co-norm

\[ I(a_1) \land I(a_2) = \max\{I(a_1) + I(a_2) - 1, 0\} \]
\[ I(a_1) \lor I(a_2) = \min\{I(a_1) + I(a_2), 1\} \]
\[ \neg I(a_1) = 1 - I(a_1) \]

rule is satisfied iff \( I(\text{body}) \leq I(\text{head}) \)
distance to satisfaction \( d(r) = \max\{0, I(\text{body}) - I(\text{head})\} \)

0.6 \quad 0.6 \quad 0.5

\text{WorksFor}(John, FBK) \land \text{BossOf}(John, Jack) \rightarrow \text{WorksFor}(Jack, FBK)
in a nutshell (2/3)

Lukasiewicz t-norm/co-norm

\[ I(a_1) \land I(a_2) = \max\{I(a_1) + I(a_2) - 1, 0\} \]
\[ I(a_1) \lor I(a_2) = \min\{I(a_1) + I(a_2), 1\} \]
\[ \neg I(a_1) = 1 - I(a_1) \]

rule is satisfied iff \( I(\text{body}) \leq I(\text{head}) \)
distance to satisfaction \( d(r) = \max\{0, I(\text{body}) - I(\text{head})\} \)

\[
\begin{array}{ccc}
0.6 & 0.6 & 0.5 \\
\text{WorksFor}(John, FBK) \land \text{BossOf}(John, Jack) \rightarrow \text{WorksFor}(Jack, FBK) & 0.8 & 0.9 & 0.3
\end{array}
\]

\( \checkmark \quad \times \)
in a nutshell (2/3)

Lukasiewicz t-norm/co-norm

\[
I(a_1) \land I(a_2) = \max\{I(a_1) + I(a_2) - 1, 0\}
\]

\[
I(a_1) \lor I(a_2) = \min\{I(a_1) + I(a_2), 1\}
\]

\[
\neg I(a_1) = 1 - I(a_1)
\]

rule is satisfied iff

\[
I(\text{body}) \leq I(\text{head})
\]

distance to satisfaction

\[
d(r) = \max\{0, I(\text{body}) - I(\text{head})\}
\]

\[
\begin{array}{ccc}
0.6 & 0.6 & 0.5 \\
\text{WorksFor}(\text{John}, FBK) \land \text{BossOf}(\text{John}, \text{Jack}) \rightarrow \text{WorksFor}(\text{Jack}, FBK) & 0.8 & 0.9 & 0.3 \checkmark \\
\end{array}
\]

\[
d(r) = 0.4
\]
in a nutshell (3/3)

\[ f(I) = \frac{1}{Z} \exp \left[ - \sum_{r \in R} w_r d(r)^p \right] \]
in a nutshell (3/3)

\[ f(I) = \frac{1}{Z} \exp \left[ - \sum_{r \in R} w_r d(r)^p \right] \]

constant
in a nutshell (3/3)

\[ f(I) = \frac{1}{Z} \exp \left[ - \sum_{r \in R} w_r d(r)^p \right] \]

constant \quad \text{all rules}
in a nutshell (3/3)

\[ f(I) = \frac{1}{Z} \exp \left[ - \sum_{r \in R} w_r d(r)^p \right] \]
in a nutshell (3/3)

\[ f(I) = \frac{1}{Z} \exp \left[ - \sum_{r \in R} w_r d(r)^p \right] \]

- constant
- weight
- distance to satisfaction
- all rules
in a nutshell (3/3)

\[ f(I) = \frac{1}{Z} \exp \left[ -\sum_{r \in R} w_r d(r)^p \right] \}

constant

weight

distance to satisfaction

all rules

\{1,2\}
Most Probable Explanation (MPE): overall interpretation with the maximum probability
NLP annotations → Classes

Classes → Annotation coherence
NLP annotations $\rightarrow$ Classes

$M$  mention
$A^T_i$  candidate annotation for task $T$ on $M$
$c$  ontological class from background knowledge $K$
Improving NLP Entity Annotations via Ontological Knowledge

Marco Rospocher

NLP annotations $\rightarrow$ Classes

$M_i$ mention

$A^T_i$ candidate annotation for task $T$ on $M$

$c$ ontological class from background knowledge $K$

$w(M, A^T_i): \text{Ann}_T(M, A^T_i) \land \text{ImpCl}_T(A^T_i, c) \rightarrow \text{ClAnn}_T(M, A^T_i, c)$
NLP annotations $\rightarrow$ Classes

$M$ mention

$A^T_i$ candidate annotation for task $T$ on $M$

$c$ ontological class from background knowledge $K$

NLP annotation

$$w(M, A^T_i) : \text{Ann}_T(M, A^T_i) \land \text{ImpCl}_T(A^T_i, c) \rightarrow \text{ClAnn}_T(M, A^T_i, c)$$
NLP annotations $\rightarrow$ Classes

$M$ mention

$A^T_i$ candidate annotation for task $T$ on $M$

$c$ ontological class from background knowledge $K$

NLP annotation

$w(M, A^T_i) : \text{Ann}_T(M, A^T_i) \land \text{ImpCl}_T(A^T_i, c) \rightarrow \text{ClAnn}_T(M, A^T_i, c)$

confidence score
Improving NLP Entity Annotations via Ontological Knowledge

Marco Rospocher

\[ M \] mention
\[ A^T_i \] candidate annotation for task \( T \) on \( M \)
\( c \) ontological class from background knowledge \( K \)

\[ w(M, A^T_i) : \text{Ann}_T(M, A^T_i) \land \text{ImpCl}_T(A^T_i, c) \rightarrow \text{ClAnn}_T(M, A^T_i, c) \]

confidence score  implied class
Improving NLP Entity Annotations via Ontological Knowledge

NLP annotations $\rightarrow$ Classes

- $M$: mention
- $A^T_i$: candidate annotation for task $T$ on $M$
- $c$: ontological class from background knowledge $K$

NLP annotation

$w(M,A^T_i) : \text{Ann}_T(M,A^T_i) \land \text{ImpCl}_T(A^T_i,c) \rightarrow \text{ClAnn}_T(M,A^T_i,c)$

Confidence score

Implied class annotation

Implied class
NLP annotations $\Rightarrow$ Classes

$M$ mention
$A^T_i$ candidate annotation for task $T$ on $M$
$c$ ontological class from background knowledge $K$

NLP annotation

$$w(M, A^T_i) : \text{Ann}_T(M, A^T_i) \wedge \text{ImpCl}_T(A^T_i, c) \rightarrow \text{ClAnn}_T(M, A^T_i, c)$$

confidence score

implied class
Improving NLP Entity Annotations via Ontological Knowledge

- **NLP annotations ➔ Classes**

  - $M$: mention
  - $A^T_i$: candidate annotation for task $T$ on $M$
  - $c$: **ontological class** from background knowledge $K$

  - **NLP annotation**
    
    $w(M, A^T_i) : Ann_T(M, A^T_i) \land \text{ImpCl}_T(A^T_i, c) \rightarrow \text{ClAnn}_T(M, A^T_i, c)$

    - **confidence score**
    - **implied class annotation**

  - **implied class**
NLP annotations $\rightarrow$ Classes

$M$ mention
$A_i^T$ candidate annotation for task $T$ on $M$
$c$ ontological class from background knowledge $K$

NLP annotation
$w(M, A_i^T): \text{Ann}_T(M, A_i^T) \land \text{ImpCl}_T(A_i^T, c) \rightarrow \text{ClAnn}_T(M, A_i^T, c)$

confidence score

implied class annotation

implied class
NLP annotations $\rightarrow$ Classes

$M_i$ mention

$A^T_i$ candidate annotation for task $T$ on $M$

$c$ ontological class from background knowledge $K$

NLP annotation

$w(M, A^T_i) : \text{Ann}_T(M, A^T_i) \land \text{ImpCl}_T(A^T_i, c) \rightarrow \text{ClAnn}_T(M, A^T_i, c)$

confidence score

implied class annotation

implied class
NLP annotations $\rightarrow$ Classes

$\text{ImpCl}_{\text{NERC}}(t, c)$
Leverage a gold standard corpus $G$ annotated with NERC types and ontological classes (or EL annotations)
Leverage a gold standard corpus $G$ annotated with NERC types and ontological classes (or EL annotations)

1.0 : $\text{Gold}_{NERC}(m,t) \land \text{ImpCl}_{NERC}(t,c) \rightarrow \text{Gold}_C(m,c)$

1.0 : $\text{Gold}_{NERC}(m,t) \land \neg\text{ImpCl}_{NERC}(t,c) \rightarrow \neg\text{Gold}_C(m,c)$
Leverage a **gold standard corpus** $G$ annotated with NERC types and ontological classes (or EL annotations)

$$1.0 : \text{Gold}_{\text{NERC}}(m,t) \land \text{ImpCl}_{\text{NERC}}(t,c) \rightarrow \text{Gold}_C(m,c)$$

$$1.0 : \text{Gold}_{\text{NERC}}(m,t) \land \neg\text{ImpCl}_{\text{NERC}}(t,c) \rightarrow \neg\text{Gold}_C(m,c)$$
Leverage a gold standard corpus $G$ annotated with NERC types and ontological classes (or EL annotations)

$$\text{ImpCl}_{\text{NERC}}(t,c)$$

1. $\text{Gold}_{\text{NERC}}(m,t) \land \text{ImpCl}_{\text{NERC}}(t,c) \rightarrow \text{Gold}_C(m,c)$
2. $\text{Gold}_{\text{NERC}}(m,t) \land \neg \text{ImpCl}_{\text{NERC}}(t,c) \rightarrow \neg \text{Gold}_C(m,c)$

$$\text{ImpCl}_{\text{EL}}(e,c)$$
Improving NLP Entity Annotations via Ontological Knowledge

Marco Rospocher

NLP annotations → Classes

ImpCl_{NERC}(t, c)

Leverage a gold standard corpus $G$ annotated with NERC types and ontological classes (or EL annotations)

1.0 : Gold_{NERC}(m, t) \land ImpCl_{NERC}(t, c) \rightarrow Gold_C(m, c)

1.0 : Gold_{NERC}(m, t) \land \neg ImpCl_{NERC}(t, c) \rightarrow \neg Gold_C(m, c)

ImpCl_{EL}(e, c)

Leverage alignments between EL Knowledge Base and Background Knowledge $K$
Improving NLP Entity Annotations via Ontological Knowledge

Marco Rospocher

Leverage a gold standard corpus $G$ annotated with NERC types and ontological classes (or EL annotations)

$$\text{ImpCl}_{NERC}(t, c)$$

Leverage alignments between EL Knowledge Base and Background Knowledge $K$

1.0 : $\text{Gold}_{NERC}(m, t) \land \text{ImpCl}_{NERC}(t, c) \rightarrow \text{Gold}_C(m, c)$

1.0 : $\text{Gold}_{NERC}(m, t) \land \neg\text{ImpCl}_{NERC}(t, c) \rightarrow \neg\text{Gold}_C(m, c)$

$$\text{ImpCl}_{EL}(e, c) \begin{cases} 1 & \text{entity } e \text{ is instance of } c \\ 0 & \text{otherwise} \end{cases}$$

Leverage alignments between EL Knowledge Base and Background Knowledge $K$
Classes $\rightarrow$ Annotation coherence

\begin{align*}
    w_1 : \text{ClAnn}_{NERC}(m,t,c) \land \text{ClAnn}_{EL}(m,e,c) & \rightarrow \text{Ann}_{PSL}(m,t,e) \\
    w_2 : \text{ClAnn}_{NERC}(m,t,c) \land \neg \text{ClAnn}_{EL}(m,e,c) & \rightarrow \neg \text{Ann}_{PSL}(m,t,e) \\
    w_3 : \neg \text{ClAnn}_{NERC}(m,t,c) \land \text{ClAnn}_{EL}(m,e,c) & \rightarrow \neg \text{Ann}_{PSL}(m,t,e)
\end{align*}
Classes $\rightarrow$ Annotation coherence

coherence estimation

$w_1 : \text{ClAnn}_{NERC}(m,t,c) \land \text{ClAnn}_{EL}(m,e,c) \rightarrow \text{Ann}_{PSL}(m,t,e)$

$w_2 : \text{ClAnn}_{NERC}(m,t,c) \land \neg\text{ClAnn}_{EL}(m,e,c) \rightarrow \neg\text{Ann}_{PSL}(m,t,e)$

$w_3 : \neg\text{ClAnn}_{NERC}(m,t,c) \land \text{ClAnn}_{EL}(m,e,c) \rightarrow \neg\text{Ann}_{PSL}(m,t,e)$
Classes $\rightarrow$ Annotation coherence

coherence estimation

\[ w_1 : \text{ClAnn}_{NERC}(m, t, c) \land \text{ClAnn}_{EL}(m, e, c) \rightarrow \text{Ann}_{PSL}(m, t, e) \]

\[ w_2 : \text{ClAnn}_{NERC}(m, t, c) \land \neg \text{ClAnn}_{EL}(m, e, c) \rightarrow \neg \text{Ann}_{PSL}(m, t, e) \]

\[ w_3 : \neg \text{ClAnn}_{NERC}(m, t, c) \land \text{ClAnn}_{EL}(m, e, c) \rightarrow \neg \text{Ann}_{PSL}(m, t, e) \]

hyperparameters
MPE Inference

• Determine soft-truth value of $\text{Ann}_{PSL}$ for all combination of annotations for a given mention

• Best combination: highest soft-truth value of $\text{Ann}_{PSL}$

• Trust model prediction only if above a given threshold
Example

*Lincoln* is based in *Michigan.*
Example

Lincoln is based in Michigan.

\[ 0.9 : \text{Ann}_{\text{NERC}}(L, \text{ORG}) \land \text{ImpCl}_{\text{NERC}}(\text{ORG}, c) \rightarrow \text{ClAnn}_{\text{NERC}}(L, \text{ORG}, c) \]

\[ 0.1 : \text{Ann}_{\text{NERC}}(L, \text{PER}) \land \text{ImpCl}_{\text{NERC}}(\text{PER}, c) \rightarrow \text{ClAnn}_{\text{NERC}}(L, \text{PER}, c) \]
Example

\textbf{Lincoln} is based in Michigan.

0.9 : \text{Ann}_{\text{NERC}}(L, \text{ORG}) \land \text{ImpCl}_{\text{NERC}}(\text{ORG}, c) \rightarrow \text{ClAnn}_{\text{NERC}}(L, \text{ORG}, c)

0.1 : \text{Ann}_{\text{NERC}}(L, \text{PER}) \land \text{ImpCl}_{\text{NERC}}(\text{PER}, c) \rightarrow \text{ClAnn}_{\text{NERC}}(L, \text{PER}, c)

0.5 : \text{Ann}_{\text{EL}}(L, \text{A. Lincoln}) \land \text{ImpCl}_{\text{EL}}(\text{A. Lincoln}, c) \rightarrow \text{ClAnn}_{\text{EL}}(L, \text{A. Lincoln}, c)

0.3 : \text{Ann}_{\text{EL}}(L, \text{Lincoln MC}) \land \text{ImpCl}_{\text{EL}}(\text{Lincoln MC}, c) \rightarrow \text{ClAnn}_{\text{EL}}(L, \text{Lincoln MC}, c)

0.2 : \text{Ann}_{\text{EL}}(L, \text{Lincoln UK}) \land \text{ImpCl}_{\text{EL}}(\text{Lincoln UK}, c) \rightarrow \text{ClAnn}_{\text{EL}}(L, \text{Lincoln UK}, c)
Example

Lincoln is based in Michigan.

0.9 : $\text{Ann}_{\text{NERC}}(L, \text{ORG}) \land \text{ImpCl}_{\text{NERC}}(\text{ORG}, c) \rightarrow \text{ClAnn}_{\text{NERC}}(L, \text{ORG}, c)$

0.1 : $\text{Ann}_{\text{NERC}}(L, \text{PER}) \land \text{ImpCl}_{\text{NERC}}(\text{PER}, c) \rightarrow \text{ClAnn}_{\text{NERC}}(L, \text{PER}, c)$

0.5 : $\text{Ann}_{\text{EL}}(L, \text{A. Lincoln}) \land \text{ImpCl}_{\text{EL}}(\text{A. Lincoln}, c) \rightarrow \text{ClAnn}_{\text{EL}}(L, \text{A. Lincoln}, c)$

0.3 : $\text{Ann}_{\text{EL}}(L, \text{Lincoln MC}) \land \text{ImpCl}_{\text{EL}}(\text{Lincoln MC}, c) \rightarrow \text{ClAnn}_{\text{EL}}(L, \text{Lincoln MC}, c)$

0.2 : $\text{Ann}_{\text{EL}}(L, \text{Lincoln UK}) \land \text{ImpCl}_{\text{EL}}(\text{Lincoln UK}, c) \rightarrow \text{ClAnn}_{\text{EL}}(L, \text{Lincoln UK}, c)$

10 : $\text{ClAnn}_{\text{NERC}}(m, t, c) \land \text{ClAnn}_{\text{EL}}(m, e, c) \rightarrow \text{Ann}_{\text{PSL}}(m, t, e)$

10 : $\text{ClAnn}_{\text{NERC}}(m, t, c) \land \neg \text{ClAnn}_{\text{EL}}(m, e, c) \rightarrow \neg \text{Ann}_{\text{PSL}}(m, t, e)$

10 : $\neg \text{ClAnn}_{\text{NERC}}(m, t, c) \land \text{ClAnn}_{\text{EL}}(m, e, c) \rightarrow \neg \text{Ann}_{\text{PSL}}(m, t, e)$
Application and Evaluation
Background Knowledge

6,016,695 entities
Taxonomy of 568,255 classes

[Suchanek et al., 2007]
Tools

- **NERC**: Stanford CoreNLP [Finkel et al., 2005]

- **EL**: DBpedia Spotlight [Daiber et al., 2013]
NERC+EL Datasets

- AIDA CoNLL-YAGO [Hoffart et al., 2011]
- MEANTIME [Minard et al., 2016]
- TAC-KBP [Ji et al., 2011]
NERC+EL Datasets

- AIDA CoNLL-YAGO [Hoffart et al., 2011]
  \[ \text{ImpCl}_{\text{NERC}} \] learned from AIDA CoNLL-YAGO (train)

- MEANTIME [Minard et al., 2016]

- TAC-KBP [Ji et al., 2011]
<table>
<thead>
<tr>
<th>PER (4522)</th>
<th>ORG (4564)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PhysicalEntity100001930 (.991)</td>
<td>YagoPermanentlyLocatedEntity (.945)</td>
</tr>
<tr>
<td>CausalAgent100007347 (.988)</td>
<td>Abstraction100002137 (.945)</td>
</tr>
<tr>
<td>Object100002684 (.963)</td>
<td>YagoLegalActorGeo (.938)</td>
</tr>
<tr>
<td>YagoLegalActorGeo (.963)</td>
<td>YagoLegalActor (.925)</td>
</tr>
<tr>
<td>Whole100003553 (.962)</td>
<td>Group100031264 (.924)</td>
</tr>
<tr>
<td>YagoLegalActor (.961)</td>
<td>SocialGroup107950920 (.923)</td>
</tr>
<tr>
<td>LivingThing100004258 (.960)</td>
<td>Organization108008335 (.914)</td>
</tr>
<tr>
<td>Organism100004475 (.960)</td>
<td>Association108049401 (.642)</td>
</tr>
<tr>
<td>Person100007846 (.960)</td>
<td>Club108227214 (.637)</td>
</tr>
<tr>
<td>WikicatLivingPeople (.850)</td>
<td>Unit108189659 (.340)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LOC (6689)</th>
<th>MISC (2764)</th>
</tr>
</thead>
<tbody>
<tr>
<td>YagoPermanentlyLocatedEntity (.986)</td>
<td>YagoPermanentlyLocatedEntity (.843)</td>
</tr>
<tr>
<td>YagoLegalActorGeo (.967)</td>
<td>YagoLegalActorGeo (.679)</td>
</tr>
<tr>
<td>PhysicalEntity100001930 (.909)</td>
<td>PhysicalEntity100001930 (.614)</td>
</tr>
<tr>
<td>Object100002684 (.907)</td>
<td>Object100002684 (.609)</td>
</tr>
<tr>
<td>YagoGeoEntity (.905)</td>
<td>YagoGeoEntity (.591)</td>
</tr>
<tr>
<td>Location100027167 (.889)</td>
<td>Location100027167 (.572)</td>
</tr>
<tr>
<td>Region108630985 (.883)</td>
<td>Region108630985 (.571)</td>
</tr>
<tr>
<td>District108552138 (.866)</td>
<td>AdministrativeDistrict108491826 (.568)</td>
</tr>
<tr>
<td>AdministrativeDistrict108491826 (.865)</td>
<td>District108552138 (.568)</td>
</tr>
<tr>
<td>Country108544813 (.524)</td>
<td>Country108544813 (.549)</td>
</tr>
</tbody>
</table>
Research Question

Does the ontology-driven PSL4EA a posteriori joint revision of the annotations from Stanford NER and DBpedia Spotlight, improve their NERC and EL performances?
Research Question

Does the ontology-driven PSL4EA a posteriori joint revision of the annotations from Stanford NER and DBpedia Spotlight, improve their NERC and EL performances?
## Results

<table>
<thead>
<tr>
<th></th>
<th>type</th>
<th>link</th>
<th>type+link</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( P )</td>
<td>( R )</td>
<td>( F_1 )</td>
</tr>
<tr>
<td><strong>AIDA (5616)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>standard</td>
<td>.943</td>
<td>.875</td>
<td>.908</td>
</tr>
<tr>
<td>with PSL4EA</td>
<td>.947</td>
<td>.879</td>
<td>.912</td>
</tr>
<tr>
<td>( \Delta )</td>
<td>.004</td>
<td>.004</td>
<td>.004</td>
</tr>
<tr>
<td><strong>MEANTIME (792)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>standard</td>
<td>.882</td>
<td>.695</td>
<td>.777</td>
</tr>
<tr>
<td>with PSL4EA</td>
<td>.902</td>
<td>.711</td>
<td>.795</td>
</tr>
<tr>
<td>( \Delta )</td>
<td>.020</td>
<td>.016</td>
<td>.018</td>
</tr>
<tr>
<td><strong>TAC-KBP (4969)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>standard</td>
<td>.911</td>
<td>.652</td>
<td>.760</td>
</tr>
<tr>
<td>with PSL4EA</td>
<td>.925</td>
<td>.662</td>
<td>.772</td>
</tr>
<tr>
<td>( \Delta )</td>
<td>.014</td>
<td>.010</td>
<td>.012</td>
</tr>
</tbody>
</table>

*bold = statistical significant (approx. rand. test)*
## Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Type</th>
<th>Link</th>
<th>Type+Link</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P$</td>
<td>$R$</td>
<td>$F_1$</td>
</tr>
<tr>
<td><strong>AIDA (5616)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>.943</td>
<td>.875</td>
<td>.908</td>
</tr>
<tr>
<td>With PSL4EA</td>
<td>.947</td>
<td>.879</td>
<td>.912</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>.004</td>
<td>.004</td>
<td>.004</td>
</tr>
<tr>
<td><strong>MEANTIME (792)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>.882</td>
<td>.695</td>
<td>.777</td>
</tr>
<tr>
<td>With PSL4EA</td>
<td>.902</td>
<td>.711</td>
<td>.795</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>.020</td>
<td>.016</td>
<td>.018</td>
</tr>
<tr>
<td><strong>TAC-KBP (4969)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>.911</td>
<td>.652</td>
<td>.760</td>
</tr>
<tr>
<td>With PSL4EA</td>
<td>.925</td>
<td>.662</td>
<td>.772</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>.014</td>
<td>.010</td>
<td>.012</td>
</tr>
</tbody>
</table>

*bold = statistical significant (approx. rand. test)*
## Results

<table>
<thead>
<tr>
<th></th>
<th>type</th>
<th></th>
<th>link</th>
<th></th>
<th>type+link</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P$</td>
<td>$R$</td>
<td>$F_1$</td>
<td>$P$</td>
<td>$R$</td>
<td>$F_1$</td>
</tr>
<tr>
<td>AIDA (5616)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>standard</td>
<td>.943</td>
<td>.875</td>
<td>.908</td>
<td>.662</td>
<td>.652</td>
<td>.656</td>
</tr>
<tr>
<td>with PSL4EA</td>
<td>.947</td>
<td>.879</td>
<td>.912</td>
<td>.670</td>
<td>.659</td>
<td>.665</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>.004</td>
<td>.004</td>
<td>.004</td>
<td>.008</td>
<td>.007</td>
<td>.009</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>.012</td>
<td>.010</td>
<td>.010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEANETIME (792)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>standard</td>
<td>.882</td>
<td>.695</td>
<td>.777</td>
<td>.703</td>
<td>.556</td>
<td>.621</td>
</tr>
<tr>
<td>with PSL4EA</td>
<td>.902</td>
<td>.711</td>
<td>.795</td>
<td>.714</td>
<td>.564</td>
<td>.630</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>.020</td>
<td>.016</td>
<td>.018</td>
<td>.011</td>
<td>.008</td>
<td>.009</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>.032</td>
<td>.025</td>
<td>.028</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAC-KBP (4969)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>standard</td>
<td>.911</td>
<td>.652</td>
<td>.760</td>
<td>.401</td>
<td>.423</td>
<td>.412</td>
</tr>
<tr>
<td>with PSL4EA</td>
<td>.925</td>
<td>.662</td>
<td>.772</td>
<td>.408</td>
<td>.430</td>
<td>.419</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>.014</td>
<td>.010</td>
<td>.012</td>
<td>.007</td>
<td>.007</td>
<td>.007</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>.017</td>
<td>.018</td>
<td>.018</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*bold = statistical significant (approx. rand. test)*
Research Question

Does the ontology-driven PSL4EA a posteriori joint revision of the annotations from Stanford NER and DBpedia Spotlight, improve their NERC and EL performances?
Research Question

Does the ontology-driven PSL4EA a posteriori joint revision of the annotations from Stanford NER and DBpedia Spotlight, improve their NERC and EL performances?
Improving NLP Entity Annotations via Ontological Knowledge
<table>
<thead>
<tr>
<th>Dataset</th>
<th>Type</th>
<th>Link</th>
<th>Type+Link</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>type</td>
<td>link</td>
<td>type+link</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F&lt;sub&gt;1&lt;/sub&gt;</td>
</tr>
<tr>
<td>AIDA (5616)</td>
<td>.007 .006 .006</td>
<td>.009 .002 .006</td>
<td>.021 .012 .016</td>
</tr>
<tr>
<td></td>
<td>with JPARK</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>with PSL4EA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEANTIME (792)</td>
<td>.032 .025 .028</td>
<td>.002 .001 .001</td>
<td>.035 .028 .031</td>
</tr>
<tr>
<td></td>
<td>with JPARK</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>with PSL4EA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAC-KBP (4969)</td>
<td>.015 .011 .012</td>
<td>.011 .003 .007</td>
<td>.022 .016 .019</td>
</tr>
<tr>
<td></td>
<td>with JPARK</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>with PSL4EA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dataset</td>
<td>Model</td>
<td>type</td>
<td>link</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$P$</td>
<td>$R$</td>
</tr>
<tr>
<td>AIDA (5616)</td>
<td>with JPARK</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>with PSL4EA</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td>MEANTIME (792)</td>
<td>with JPARK</td>
<td>0.032</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>with PSL4EA</td>
<td>0.020</td>
<td>0.016</td>
</tr>
<tr>
<td>TAC-KBP (4969)</td>
<td>with JPARK</td>
<td>0.015</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>with PSL4EA</td>
<td>0.014</td>
<td>0.010</td>
</tr>
</tbody>
</table>

✓ very fast
✓ simple model construction
<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>type</th>
<th>link</th>
<th>type+link</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$P$</td>
<td>$R$</td>
<td>$F_1$</td>
</tr>
<tr>
<td>AIDA (5616)</td>
<td>with JPARK</td>
<td>.007</td>
<td>.006</td>
<td>.006</td>
</tr>
<tr>
<td></td>
<td>with PSL4EA</td>
<td>.004</td>
<td>.004</td>
<td>.004</td>
</tr>
<tr>
<td>MEANTIME (792)</td>
<td>with JPARK</td>
<td>.032</td>
<td>.025</td>
<td>.028</td>
</tr>
<tr>
<td></td>
<td>with PSL4EA</td>
<td>.020</td>
<td>.016</td>
<td>.018</td>
</tr>
<tr>
<td>TAC-KBP (4969)</td>
<td>with JPARK</td>
<td>.015</td>
<td>.011</td>
<td>.012</td>
</tr>
<tr>
<td></td>
<td>with PSL4EA</td>
<td>.014</td>
<td>.010</td>
<td>.012</td>
</tr>
</tbody>
</table>

- JPARK: very fast
- PSL4EA: intuitive formulation
- simple model construction
- extensible to cross-mention information
Conclusions

• Ontological knowledge does really help improving NLP entity annotations

• Two approaches: JPARK and PSL EA

• Instantiation of the models for the NERC and EL tasks
Conclusions

• **Empirical confirmation** (3 datasets) of the capability of the models to improve the quality of the annotations

• Applicable to “any” NERC and EL tools

• **Future Work:**
  - application to other tasks (e.g., SRL)
  - application to fine-grained NERC
  - Testing different background knowledge (e.g., DBpedia, Wikidata)
  - cross-mention coherence