Joint Posterior Revision of NLP Entity Annotations via Ontological Knowledge

Marco Rospocher

(Joint work with Francesco Corcoglioniti @ UniBZ)

UNIVERSITÀ di VERONA

Dipartimento di LINGUE E LETTERATURE STRANIERE

Dipartimento di Eccellenza 2018/2022 MIUR

MTL NLP Consortium meetup — 22.01.2021

Research Context

- Knowledge graph extraction from text

Kia has hired Peter Schreyer as chief design officer.

Research Context

- Knowledge graph extraction from text

$\boxed{\text{Kia}}$ has hired Peter Schreyer as chief design officer.

Research Context

- Knowledge graph extraction from text

Organization

$\boxed{\text{Kia}}$ has hired Peter Schreyer as chief design officer.

NLP Tasks:

- Named Entity Recognition and Classification (NERC)

Research Context

- Knowledge graph extraction from text

Organization   dbpedia:Kia_Motors

Kia has hired Peter Schreyer as chief design officer.

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

Research Context

- Knowledge graph extraction from text

Organization

dbpedia:Kia_Motors

Kia has hired Peter Schreyer as chief design officer.

framenet:employer

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

…

Research Context

- Knowledge graph extraction from text

PIKES

Pikes is a Knowledge Extraction Suite
https://pikes.fbk.eu/

Organization

dbpedia:Kia_Motors

$\boxed{\text{Kia}}$ has hired Peter Schreyer as chief design officer.

framenet:employer

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

…

What may happen in practice…

Eric Clapton is one of the greatest guitar players.

Mr. Washington was runner-up at Wimbledon in 1996.

San Jose is one of the strongest hockey teams.

What may happen in practice…

Stanford CoreNLP

DBpedia Spotlight

Eric Clapton is one of the greatest guitar players.

Mr. Washington was runner-up at Wimbledon in 1996.

San Jose is one of the strongest hockey teams.

What may happen in practice…

Person

Stanford CoreNLP

🙂 Eric Clapton is one of the greatest guitar players.

DBpedia Spotlight

dbpedia:Eric_Clapton

Mr. Washington was runner-up at Wimbledon in 1996.

San Jose is one of the strongest hockey teams.

What may happen in practice...

Stanford CoreNLP

DBpedia Spotlight

Person

Eric Clapton is one of the greatest guitar players.

dbpedia:Eric_Clapton

Person

Mr. Washington was runner-up at Wimbledon in 1996.

dbpedia:Washington_(state)

San Jose is one of the strongest hockey teams.

What may happen in practice…

Stanford CoreNLP

DBpedia Spotlight

Person

Eric Clapton is one of the greatest guitar players.

dbpedia:Eric_Clapton

Person

Mr. Washington was runner-up at Wimbledon in 1996.

dbpedia:Washington_(state)

Location

San Jose is one of the strongest hockey teams.

dbpedia:San_Jose_Sharks

What may happen in practice…

Stanford CoreNLP

DBpedia Spotlight

Person

Eric Clapton is one of the greatest guitar players.

dbpedia:Eric_Clapton

Person

Mr. Washington was runner-up at Wimbledon in 1996.

dbpedia:Washington_(state)

Location | Organization | Person | Misc

San Jose is one of the strongest hockey teams.

dbpedia:San_Jose_Sharks | dbpedia:San_Jose,_California | dbpedia:San_Jose_Earthquakes | dbpedia:SAP_Cent

What may happen in practice…

Stanford CoreNLP

DBpedia Spotlight

Person

Eric Clapton is one of the greatest guitar players.

dbpedia:Eric_Clapton

Person

Mr. Washington was runner-up at Wimbledon in 1996.

dbpedia:Washington_(state)

Organization

San Jose is one of the strongest hockey teams.

dbpedia:San_Jose_Sharks

How can we improve the coherence of the various NLP annotations on an entity mention?

Leveraging Ontological Knowledge!

… $token_1$ $token_2$ $token_3$ $token_4$ $token_5$ $token_6$ ….

Leveraging Ontological Knowledge!

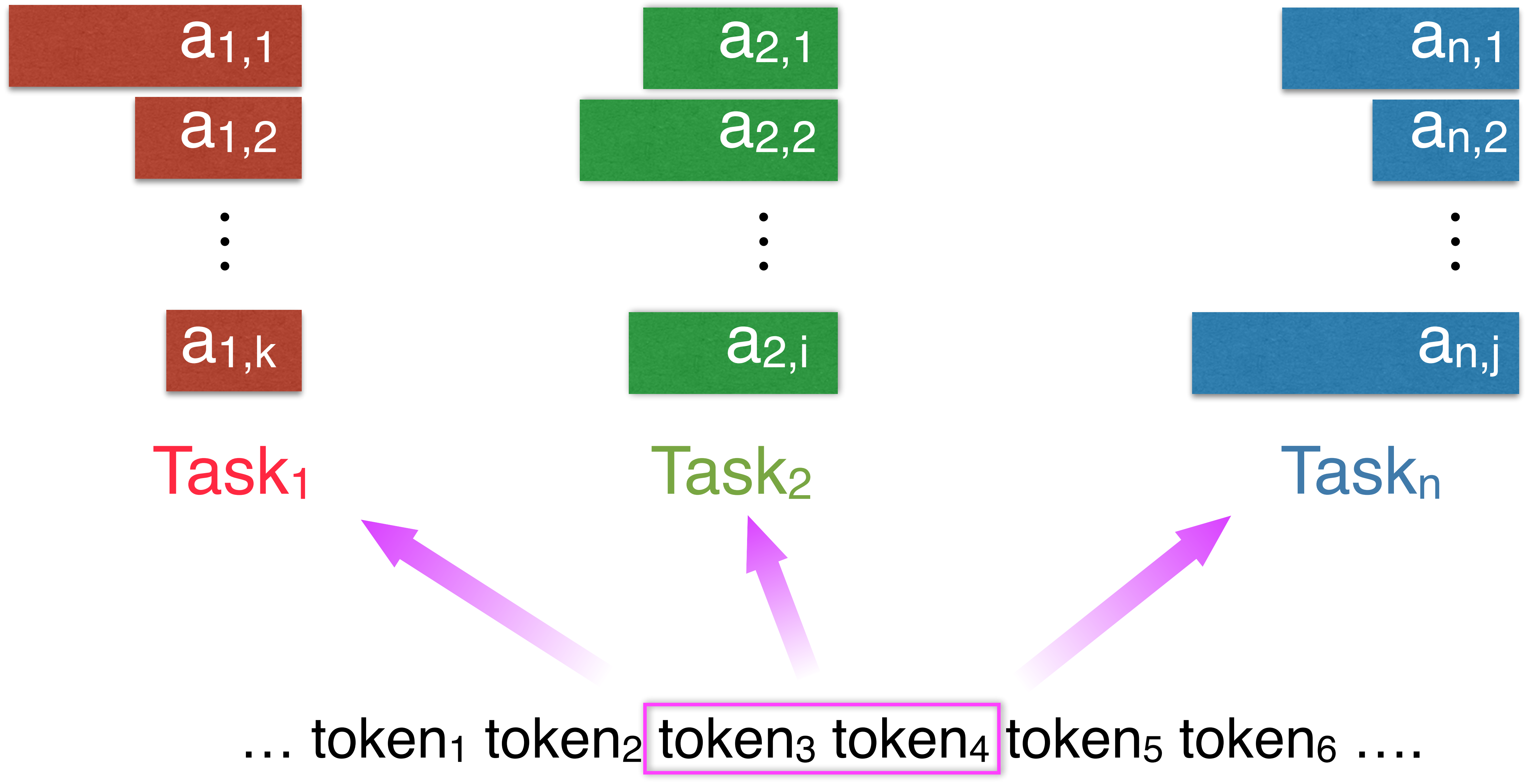
… $token_1$ $token_2$ $\boxed{token_3\ token_4}$ $token_5$ $token_6$ ….

Leveraging Ontological Knowledge!

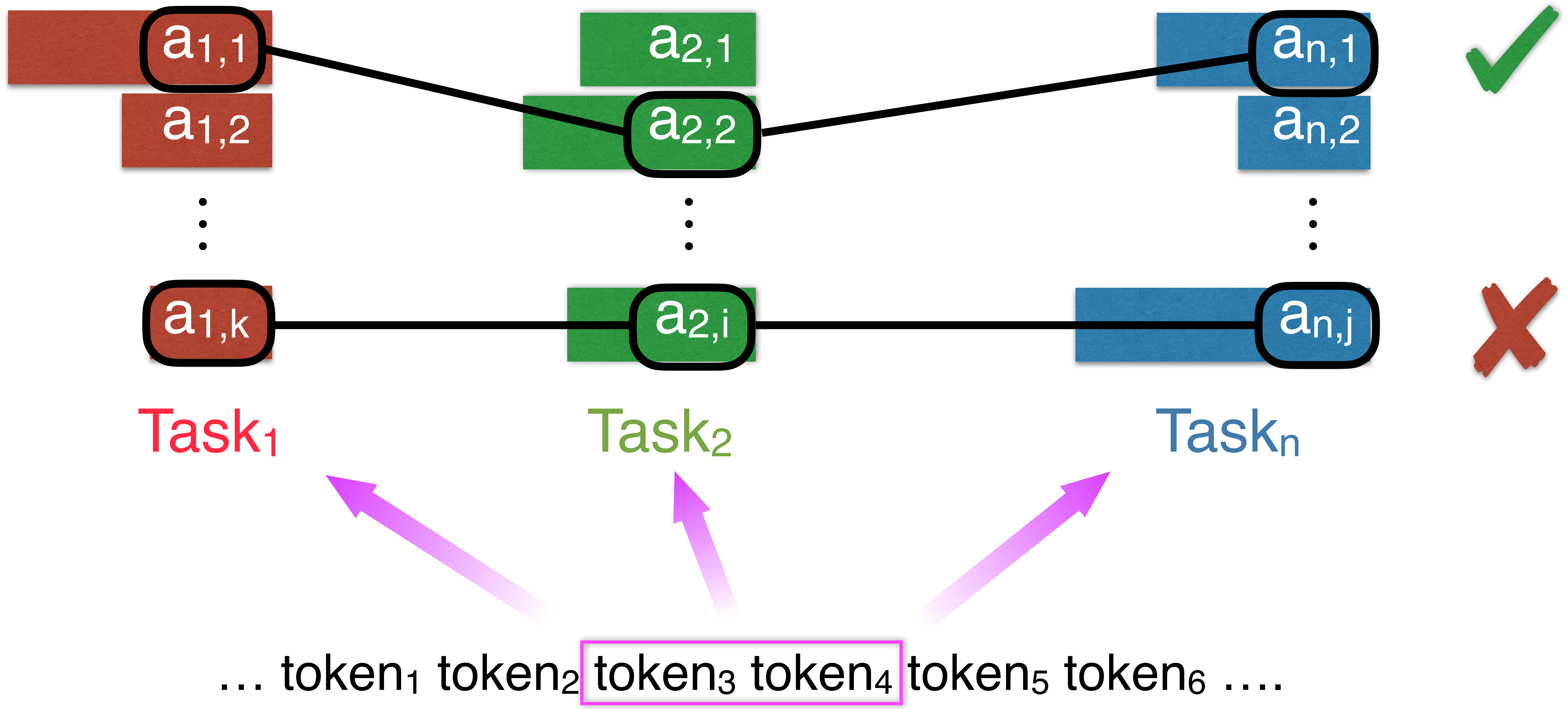
Task_1 Task_2 Task_n

… token_1 token_2 $\boxed{\text{token}_3\ \text{token}_4}$ token_5 token_6 ….

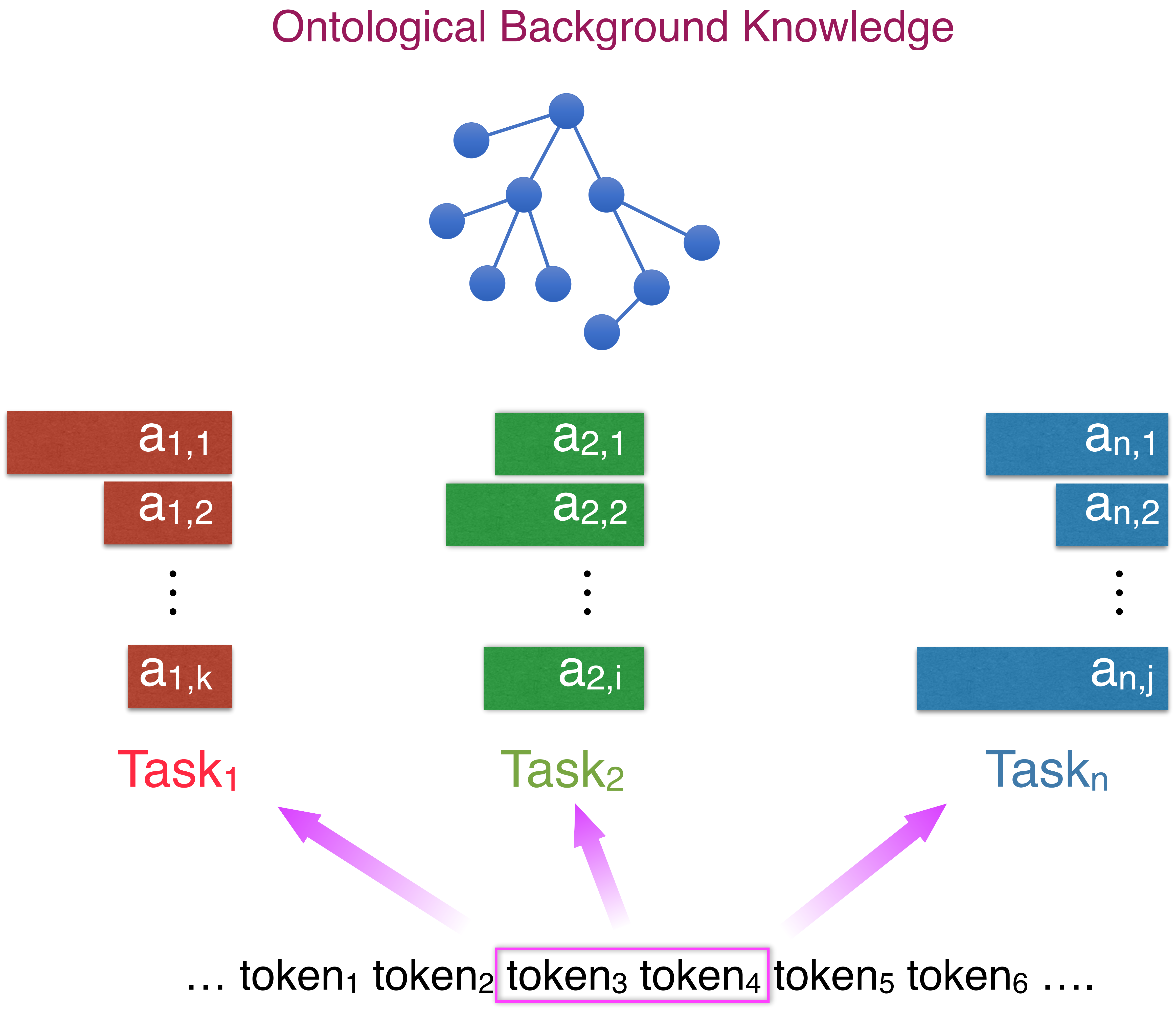
Leveraging Ontological Knowledge!



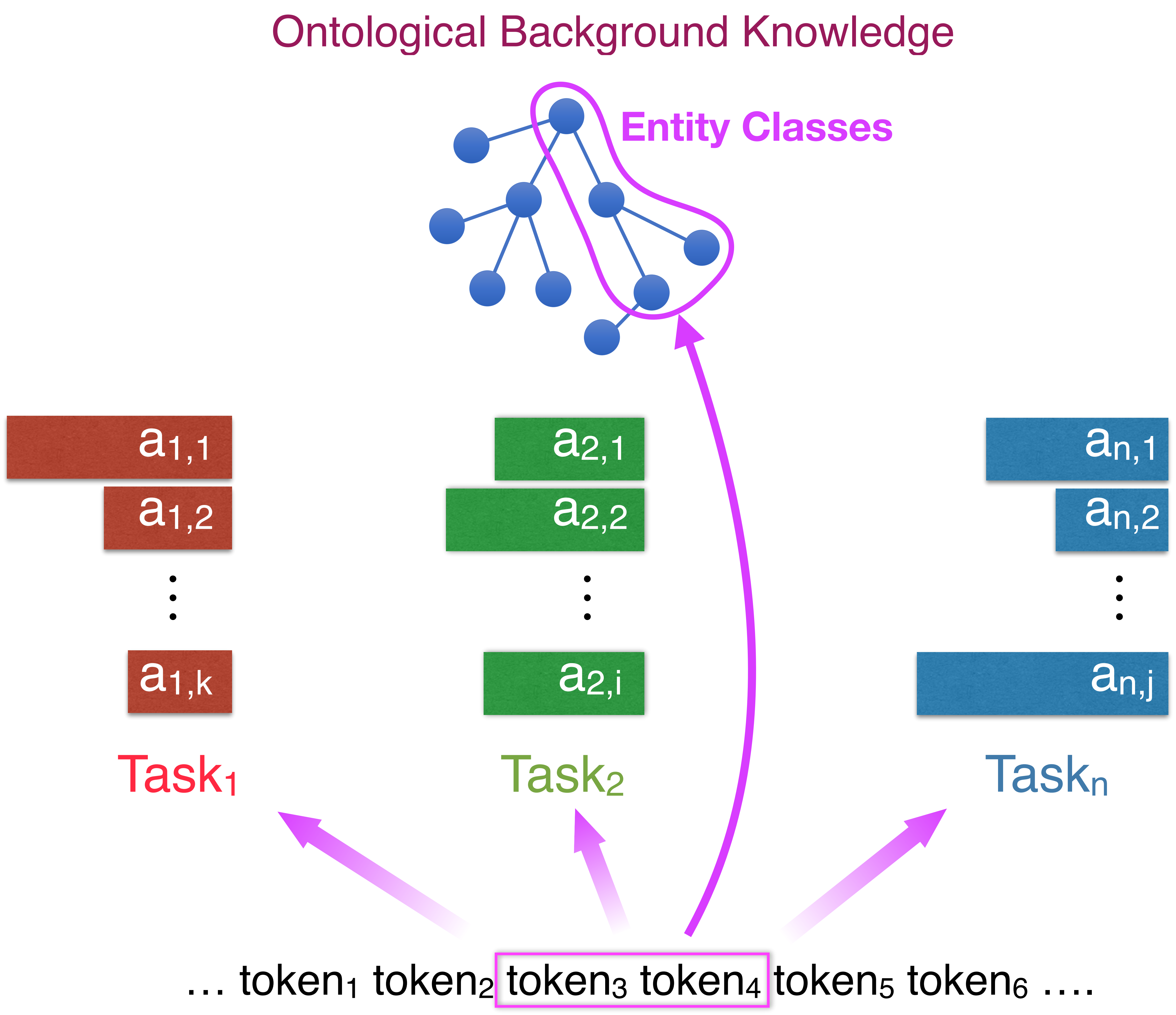
Leveraging Ontological Knowledge!



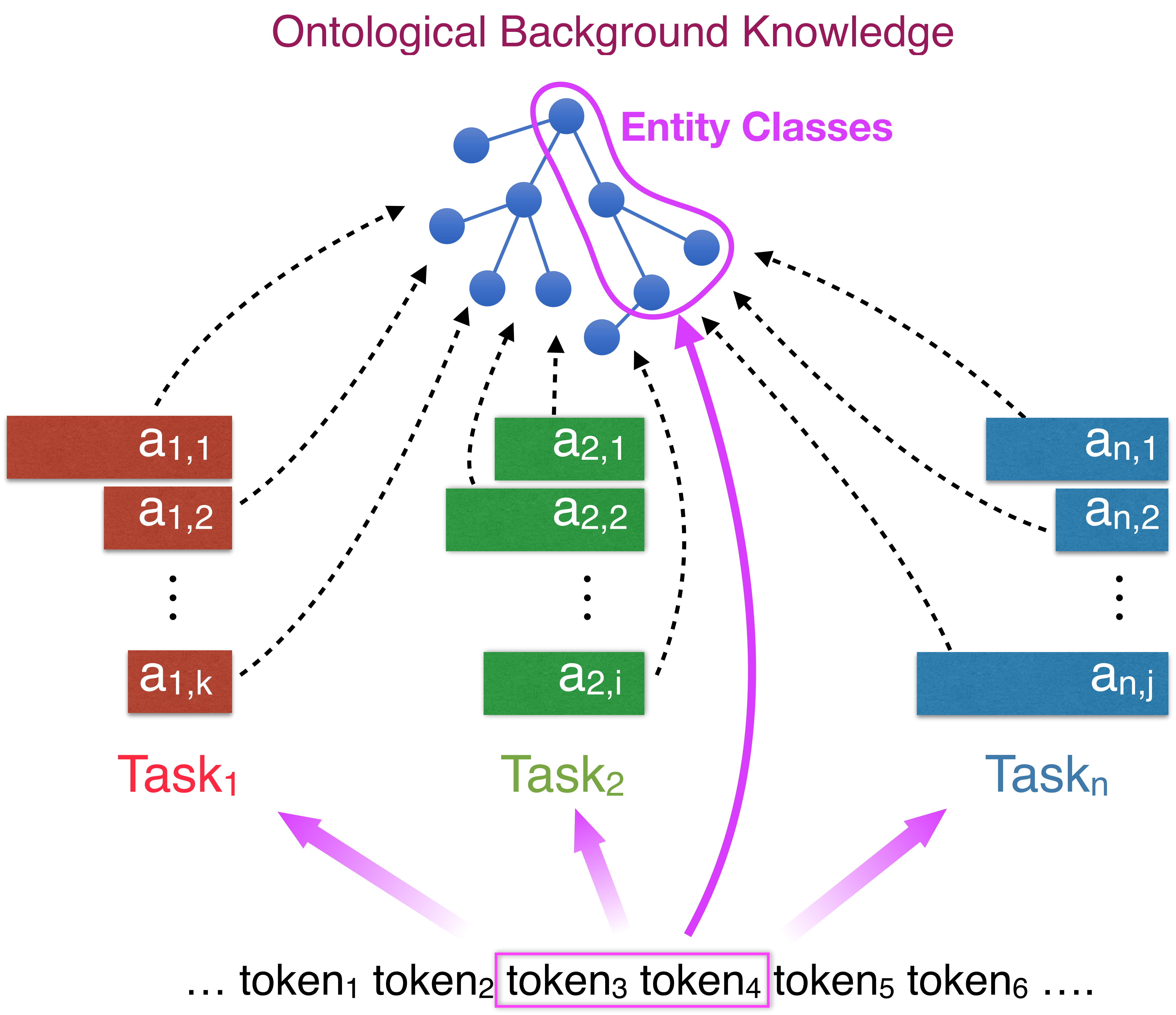
Leveraging Ontological Knowledge!



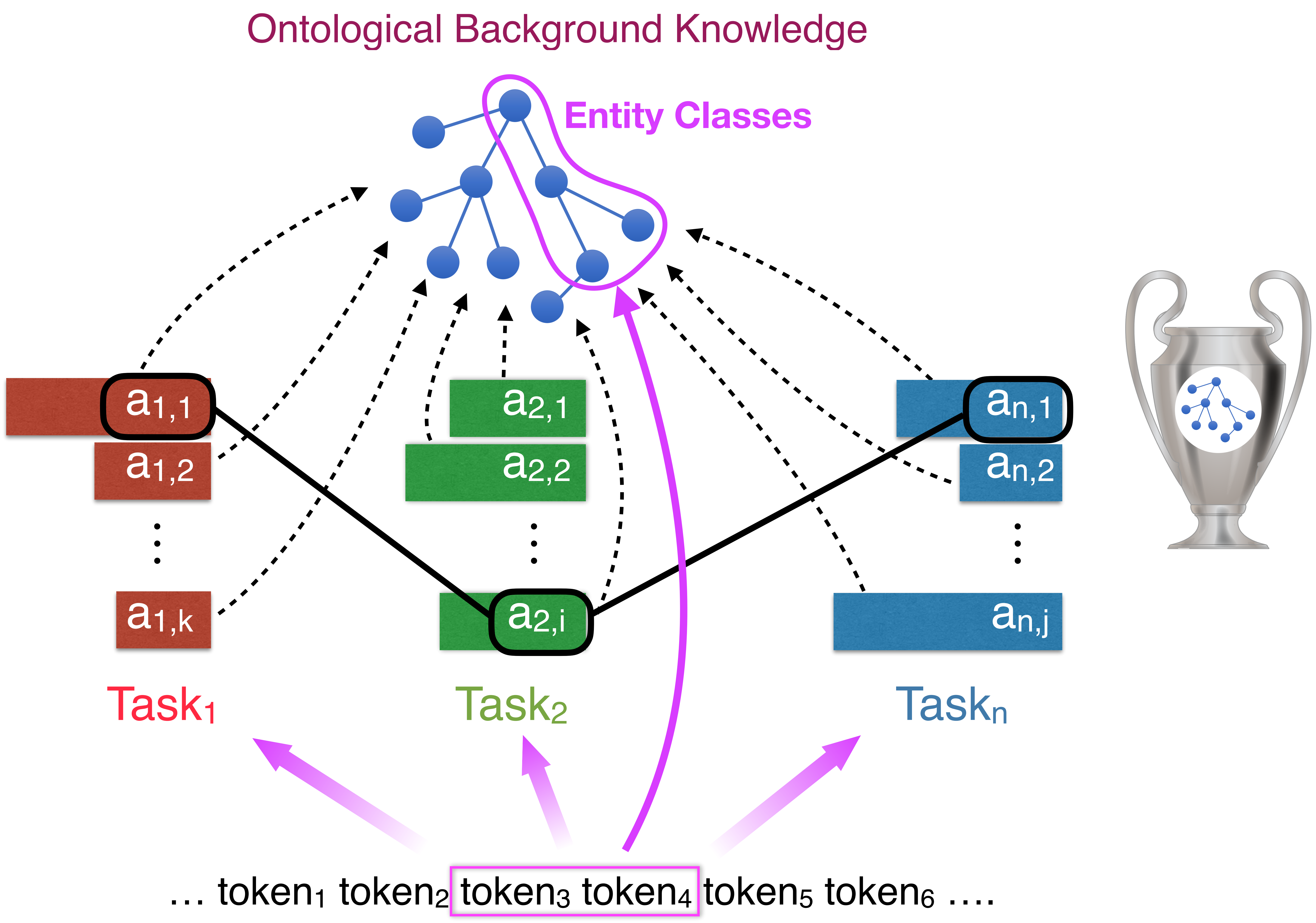
Leveraging Ontological Knowledge!



Leveraging Ontological Knowledge!



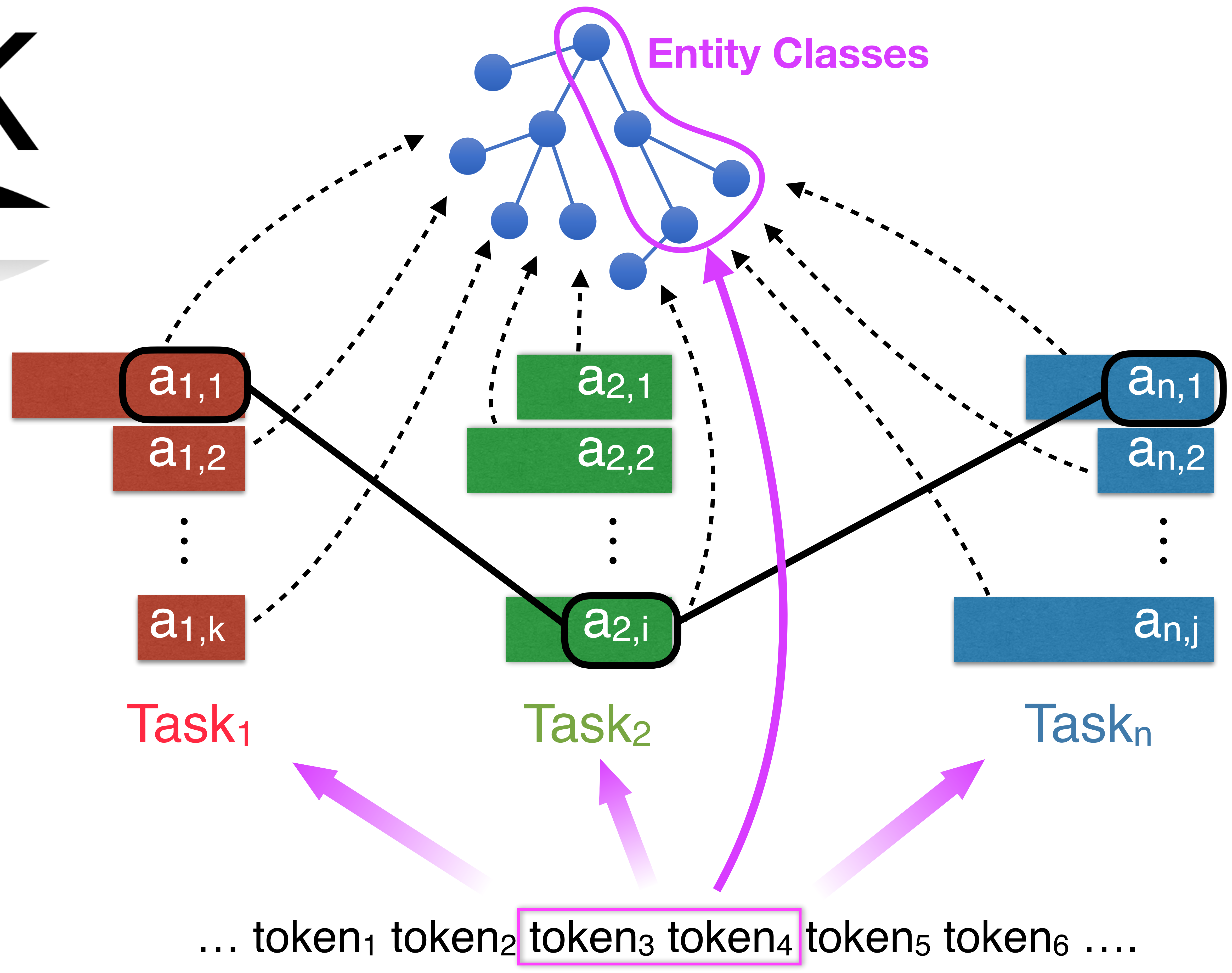
Leveraging Ontological Knowledge!



Leveraging Ontological Knowledge!

JPARK

Ontological Background Knowledge



How does JPARK work?

$$\mathrm{P}\left(\boldsymbol{a} \,\middle|\, m, B, K\right)$$

How does JPARK work?

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

Entity mention

NLP Background Knowledge

$$\mathrm{P}\left(\boldsymbol{a} \,\middle|\, m, B, K\right)$$ "The" Ontological Knowledge

How does JPARK work?

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

Entity mention

NLP Background Knowledge

$\mathrm{P}\left(\boldsymbol{a} \,|\, m, B, K\right)$ "The" Ontological Knowledge

set of classes from K

$\mathrm{P}\left(\boldsymbol{a}, C \,|\, m, B, K\right)$

How does JPARK work?

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

Entity mention

NLP Background Knowledge

$\mathrm{P}\,(\boldsymbol{a}\,|\,m,\,B,\,K\,)$ "The" Ontological Knowledge

set of classes from K

$\mathrm{P}\,(\boldsymbol{a}\,,C\,|\,m,\,B,\,K\,)$

$\mathrm{P}\,(a_i\,|\,m,\,B\,)$

$\mathrm{P}\,(C\,|\,a_i\,,\,K\,)$

UNIVERSITÀ di VERONA
Dipartimento di LINGUE E LETTERATURE STRANIERE
Dipartimento di Eccellenza 2018/2022 MIUR
Marco Rospocher

How does JPARK work?

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

Entity mention

NLP Background Knowledge

$\mathrm{P}\left(\boldsymbol{a} \mid m, B, K\right)$ "The" Ontological Knowledge

set of classes from K

$\mathrm{P}\left(\boldsymbol{a}, C \mid m, B, K\right)$

confidence score

$\boxed{\mathrm{P}\left(a_i \mid m, B\right)}$

$\mathrm{P}\left(C \mid a_i, K\right)$

How does JPARK work?

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

Entity mention

NLP Background Knowledge

$P(\boldsymbol{a} \mid m, B, K)$ "The" Ontological Knowledge

set of classes from K

$P(\boldsymbol{a}, C \mid m, B, K)$

confidence score

$\boxed{P(a_i \mid m, B)}$

learned from data

$\boxed{P(C \mid a_i, K)}$

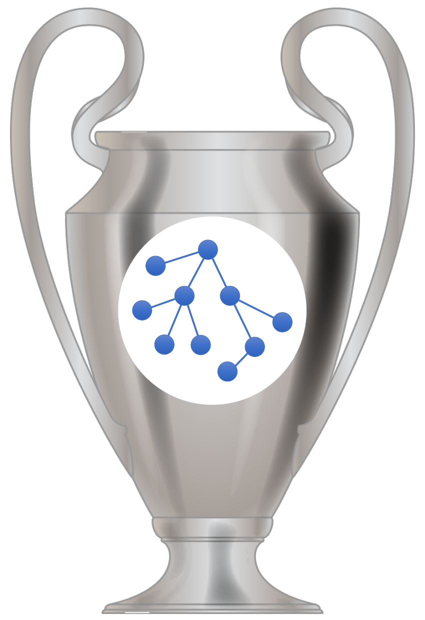
UNIVERSITÀ di VERONA
Dipartimento di LINGUE E LETTERATURE STRANIERE
Dipartimento di Eccellenza 2018/2022 MIUR

How does JPARK work?

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

Entity mention

NLP Background Knowledge



$$= \arg\max_{\boldsymbol{a}} \mathrm{P}\left(\boldsymbol{a} \,|\, m, B, K\right)$$ "The" Ontological Knowledge

set of classes from K

$$\mathrm{P}\left(\boldsymbol{a}, C \,|\, m, B, K\right)$$

confidence score

$$\boxed{\mathrm{P}\left(a_i \,|\, m, B\right)}$$

learned from data

$$\boxed{\mathrm{P}\left(C \,|\, a_i, K\right)}$$

Building a JPARK model for NERC and Entity Linking

Handling NERC Annotations

- a_{NERC} is a NERC type such as PER, ORG, LOC, MISC, …
- We have to estimate

$$\mathrm{P}\left(C \,|\, a_{\mathrm{NERC}}, K\right) =$$

- We rely on a Gold Standard G containing entity mentions that
  - are annotated with ground truth NERC types a_{NERC}
  - can be related directly or indirectly to class sets C in K

Handling NERC Annotations

- a_{NERC} is a NERC type such as PER, ORG, LOC, MISC, …
- We have to estimate

$$\mathrm{P}\left(C \,|\, a_{\mathrm{NERC}}, K\right) = \frac{\overbrace{n_G}^{\text{\# occurrences}}\left(C, a_{\mathrm{NERC}}\right)}{\sum_{C^*} n_G\left(C^*, a_{\mathrm{NERC}}\right)}$$

- We rely on a Gold Standard G containing entity mentions that
  - are annotated with ground truth NERC types a_{NERC}
  - can be related directly or indirectly to class sets C in K

Handling NERC Annotations

- a_{NERC} is a NERC type such as PER, ORG, LOC, MISC, …
- We have to estimate

occurrences

$$\mathrm{P}\left(C \,|\, a_{\mathrm{NERC}}, K\right) = (1-\alpha) \cdot \frac{n_G\left(C, a_{\mathrm{NERC}}\right)}{\sum_{C^*} n_G\left(C^*, a_{\mathrm{NERC}}\right)} + \alpha \cdot \boxed{\frac{n_K(C)}{\sum_{C^*} n_K\left(C^*\right)}}$$

Prior for unseen class sets
(e.g., class set popularity in K)

- We rely on a Gold Standard G containing entity mentions that
  - are annotated with ground truth NERC types a_{NERC}
  - can be related directly or indirectly to class sets C in K

Handling EL Annotations

- a_{EL} is an entity in the Linking knowledge base L
- The L linking knowledge base and the K ontological background knowledge are not necessary the same
  - However, in general, we can deterministically align entities/classes between the two, and so we can obtain a corresponding class set C_K for a_{EL}

$$\mathrm{P}\left(C \middle| a_{\mathrm{EL}}, K\right) =$$

Handling EL Annotations

- a_{EL} is an entity in the Linking knowledge base L
- The L linking knowledge base and the K ontological background knowledge are not necessary the same
  - However, in general, we can deterministically align entities/classes between the two, and so we can obtain a corresponding class set C_K for a_{EL}

$$\mathrm{P}\left(C \middle| a_{\mathrm{EL}}, K\right) = \begin{cases} 1 & C = C_K\left(a_{\mathrm{EL}}\right) \\ 0 & \text{otherwise} \end{cases}$$

Handling EL Annotations: NIL

- What about NIL (i.e., an entity lacking a corresponding referent in the linking knowledge base)? Can we estimate $\mathrm{P}\left(C \mid \mathrm{NIL}, K\right)$?

$$\mathrm{P}\left(C \mid \mathrm{NIL}, K\right) = \sum_{a_{\mathrm{NERC}}} \mathrm{P}\left(C \mid a_{\mathrm{NERC}}, K\right) \cdot \mathrm{P}\left(a_{\mathrm{NERC}} \mid \mathrm{NIL}, K\right)$$

Handling EL Annotations: NIL

- What about NIL (i.e., an entity lacking a corresponding referent in the linking knowledge base)? Can we estimate $\mathrm{P}\left(C \mid \mathrm{NIL}, K\right)$?

Already computed for handling NERC

$$\mathrm{P}\left(C \mid \mathrm{NIL}, K\right) = \sum_{a_{\mathrm{NERC}}} \boxed{\mathrm{P}\left(C \mid a_{\mathrm{NERC}}, K\right)} \cdot \boxed{\mathrm{P}\left(a_{\mathrm{NERC}} \mid \mathrm{NIL}, K\right)}$$

Also computable from corpus G

Choice of Ontological Classes

- Practically, we may restrict to consider a limited number of popular classes from the ontological background knowledge K
  - most of the classes in K never or rarely occur in the corpus G
  - leaf classes deep in K class taxonomy may be affected by incomplete knowledge
  - a large number of class sets slows down the use of

- Intuition: we keep only the classes in K that occur a substantial number n^* of times in G
  - By construction, this strategy ends up keeping only the top level, informative classes of K class taxonomy

Evaluation

Datasets

- Three datasets, annotated with NERC (4-type CoNLL2003) and EL (DBpedia)
  - AIDA CoNLL-YAGO [Hoffart et al., 2011] - news wire articles from Reuters
    - its train part used as corpus G
  - MEANTIME [Minard et al., 2016] - news articles from Wikinews
  - TAC-KBP [Ji et al., 2011] - news wire articles, posts from blogs, newsgroups, and fora

Dataset	Docs	Mentions					NILs					Avg. Mentions
		Total	PER	ORG	LOC	MISC	frac.	PER	ORG	LOC	MISC	per Entity
AIDA eng.train	946	23,411	.28	.27	.30	.15	.20	.44	.37	.09	.10	4.54
AIDA eng.testb	231	5,624	.29	.29	.30	.12	.20	.52	.27	.08	.14	2.91
MEANTIME	120	793	.09	.59	.32	-	.08	.15	.46	.39	-	4.42
TAC	2,231	4,969	.43	.30	.26	-	.46	.50	.32	.18	-	3.82
MERGED	2,582	11,386	.34	.32	.28	.06	.31	.50	.31	.15	.05	3.28

Background Ontological Knowledge

Version 3

# entities:	6,016,695
# classes:	568,255
# class sets:	2,126,074

Version 2016-04

# entities:	5,109,890
# classes:	754
# class sets:	413

Dump 2018.08

# entities:	5,854,808
# classes:	21,864
# class sets:	32,648

Background Ontological Knowledge

WIKIPEDIA
The Free Encyclopedia

yago
select knowledge

Version 3

# entities:	6,016,695
# classes:	568,255
# class sets:	2,126,074

DBpedia

Version 2016-04

# entities:	5,109,890
# classes:	754
# class sets:	413

WIKIDATA

Dump 2018.08

# entities:	5,854,808
# classes:	21,864
# class sets:	32,648

Background Ontological Knowledge

yago
select knowledge

Version 3
entities: 6,016,695
classes: 568,255
class sets: 2,126,074

WIKIPEDIA
The Free Encyclopedia

DBpedia

Version 2016-04
entities: 5,109,890
classes: 754
class sets: 413

WIKIDATA

Dump 2018.08
entities: 5,854,808
classes: 21,864
class sets: 32,648

owl:sameAs links

1:1

EL Knowledge Base

DBpedia

owl:sameAs links

Background Ontological Knowledge

Version 3
\# entities: 6,016,695
\# classes: 568,255
\# class sets: ~~2,126,074~~ 2,041

Version 2016-04
\# entities: 5,109,890
\# classes: 754
\# class sets: ~~413~~ 35

Dump 2018.08
\# entities: 5,854,808
\# classes: 21,864
\# class sets: ~~32,648~~ 711

owl:sameAs links

1:1

owl:sameAs links

EL Knowledge Base

Tools

Experiment A:

popular, frequently used tools

CoreNLP

[Manning et al, 2014]

CONLL2003 (PER, ORG, LOC, MISC)

DBpedia Spotlight

[Mendes et al, 2011]

DBpedia, NIL

Experiment B:

state-of-the-art neural approaches

flair

[Akbik et al, 2018]

CONLL2003 (PER, ORG, LOC, MISC)

End-to-End Neural EL

[Kolitsas et al, 2018]

Wikipedia/DBpedia, NO NIL

Does the JPARK a posteriori joint revision of annotations by NERC and EL tools, performed leveraging ontological knowledge, improve NERC and EL performance?

Evaluation Measures

- Three measures, typically adopted in NERC and EL evaluation campaigns:
  - type: a correct mention has the same span and NERC type as a gold annotation
  - link: a correct mention has the same span and EL entity as a gold annotation
  - type+link: a correct mention has the same span, NERC type, and EL entity as a gold annotation
- Metrics: precision (P), recall (R), and F_1
  - computed via TAC-KBP official scorer
- We don't specifically address the capability to detect entity mentions
  - our approach fully relies on the mentions detected by the NLP tools

Evaluation Procedure

- We create a JPARK model for each background knowledge resource (DBpedia, Yago, Wikidata)
  - AIDA eng.train (corpus G), AIDA eng.testa (hyper-parameters)
  - evaluation on: AIDA eng.testb, MEANTIME, and TAC-KBP (and MERGED)

- Experiment A (CoreNLP + DBpedia Spotlight)

baseline vs. JPARK vs. JPARK +NIL

- Experiment B (Flair + E2E Neural EL)

baseline vs. JPARK

Experiment A

CoreNLP + DBpedia Spotlight

dataset	setting	back. know.	type (F_1)	link (F_1)	type+link (F_1)
AIDA	baseline		.908	.656	.630
	JPARK	YAGO	**.006**	**.006**	**.016**
		DBpedia	**.004**	**.008**	**.014**
		Wikidata	**.008**	.005	**.015**
	m.a.i.		.053	.102	.127
MEANTIME	baseline		.777	.621	.561
	JPARK	YAGO	**.028**	.001	**.031**
		DBpedia	**.030**	-.002	**.027**
		Wikidata	**.028**	.000	**.029**
	m.a.i.		.096	.106	.162
TAC KBP	baseline		.760	.412	.376
	JPARK	YAGO	**.012**	**.007**	**.019**
		DBpedia	**.019**	**.019**	**.035**
		Wikidata	**.016**	**.017**	**.034**
	m.a.i.		.073	.172	.208
MERGED	baseline		.838	.568	.535
	JPARK	YAGO	**.011**	**.006**	**.019**
		DBpedia	**.012**	**.011**	**.024**
		Wikidata	**.013**	**.009**	**.024**
	m.a.i.		.065	.133	.165

stat. sign. results in bold

Experiment A

CoreNLP + DBpedia Spotlight

dataset	setting	back. know.	type (F_1)	+NIL	link (F_1)	+NIL	type+link (F_1)	+NIL
AIDA	baseline		.908		.656		.630	
	JPARK	YAGO	**.006**	**.005**	**.006**	**.023**	**.016**	**.035**
		DBpedia	**.004**	**.004**	**.008**	**.013**	**.014**	**.024**
		Wikidata	**.008**	**.005**	.005	**.013**	**.015**	**.027**
	m.a.i.		.053		.102		.127	
MEANTIME	baseline		.777		.621		.561	
	JPARK	YAGO	**.028**	**.021**	.001	-.005	**.031**	**.025**
		DBpedia	**.030**	**.026**	-.002	.000	**.027**	**.030**
		Wikidata	**.028**	**.024**	.000	.008	**.029**	**.042**
	m.a.i.		.096		.106		.162	
TAC KBP	baseline		.760		.412		.376	
	JPARK	YAGO	**.012**	**.013**	**.007**	**.027**	**.019**	**.049**
		DBpedia	**.019**	**.015**	**.019**	**.031**	**.035**	**.054**
		Wikidata	**.016**	**.014**	**.017**	**.026**	**.034**	**.051**
	m.a.i.		.073		.172		.208	
MERGED	baseline		.838		.568		.535	
	JPARK	YAGO	**.011**	**.009**	**.006**	**.024**	**.019**	**.041**
		DBpedia	**.012**	**.010**	**.011**	**.020**	**.024**	**.037**
		Wikidata	**.013**	**.010**	**.009**	**.019**	**.024**	**.039**
	m.a.i.		.065		.133		.165	

stat. sign. results in bold

Experiment A

CoreNLP + DBpedia Spotlight

dataset	setting	back. know.	type (F_1)	+NIL	link (F_1)	+NIL	type+link (F_1)	+NIL
AIDA	baseline		.908		.656		.630	
	JPARK	YAGO	**.006**	**.005**	**.006**	**.023**	**.016**	**.035**
		DBpedia	**.004**	**.004**	**.008**	**.013**	**.014**	**.024**
		Wikidata	**.008**	**.005**	.005	**.013**	**.015**	**.027**
	m.a.i.		.053		.102		.127	
MEANTIME	baseline		.777		.621		.561	
	JPARK	YAGO	**.028**	**.021**	.001	-.005	**.031**	**.025**
		DBpedia	**.030**	**.026**	-.002	.000	**.027**	**.030**
		Wikidata	**.028**	**.024**	.000	.008	**.029**	**.042**
	m.a.i.		.096		.106		.162	
TAC KBP	baseline		.760		.412		.376	
	JPARK	YAGO	**.012**	**.013**	**.007**	**.027**	**.019**	**.049**
		DBpedia	**.019**	**.015**	**.019**	**.031**	**.035**	**.054**
		Wikidata	**.016**	**.014**	**.017**	**.026**	**.034**	**.051**
	m.a.i.		.073		.172		.208	
MERGED	baseline		.838		.568		.535	
	JPARK	YAGO	**.011**	**.009**	**.006**	**.024**	**.019**	**.041**
		DBpedia	**.012**	**.010**	**.011**	**.020**	**.024**	**.037**
		Wikidata	**.013**	**.010**	**.009**	**.019**	**.024**	**.039**
	m.a.i.		.065		.133		.165	

stat. sign. results in bold

- Improvement (mostly stat. sign.) almost everywhere

Experiment A

CoreNLP + DBpedia Spotlight

dataset	setting	back. know.	type (F_1)	+NIL	link (F_1)	+NIL	type+link (F_1)	+NIL
AIDA	baseline		.908		.656		.630	
	JPARK	YAGO	**.006**	**.005**	**.006**	**.023**	**.016**	**.035**
		DBpedia	**.004**	**.004**	**.008**	**.013**	**.014**	**.024**
		Wikidata	**.008**	**.005**	.005	**.013**	**.015**	**.027**
	m.a.i.		.053		.102		.127	
MEANTIME	baseline		.777		.621		.561	
	JPARK	YAGO	**.028**	**.021**	.001	-.005	**.031**	**.025**
		DBpedia	**.030**	**.026**	-.002	.000	**.027**	**.030**
		Wikidata	**.028**	**.024**	.000	.008	**.029**	**.042**
	m.a.i.		.096		.106		.162	
TAC KBP	baseline		.760		.412		.376	
	JPARK	YAGO	**.012**	**.013**	**.007**	**.027**	**.019**	**.049**
		DBpedia	**.019**	**.015**	**.019**	**.031**	**.035**	**.054**
		Wikidata	**.016**	**.014**	**.017**	**.026**	**.034**	**.051**
	m.a.i.		.073		.172		.208	
MERGED	baseline		.838		.568		.535	
	JPARK	YAGO	**.011**	**.009**	**.006**	**.024**	**.019**	**.041**
		DBpedia	**.012**	**.010**	**.011**	**.020**	**.024**	**.037**
		Wikidata	**.013**	**.010**	**.009**	**.019**	**.024**	**.039**
	m.a.i.		.065		.133		.165	

stat. sign. results in bold

- Improvement (mostly stat. sign.) almost everywhere
- Particularly effective in selecting the correct annotation couple

Experiment A

CoreNLP + DBpedia Spotlight

dataset	setting	back. know.	type (F_1)	+NIL	link (F_1)	+NIL	type+link (F_1)	+NIL
AIDA	baseline		.908		.656		.630	
	JPARK	YAGO	**.006**	**.005**	**.006**	**.023**	**.016**	**.035**
		DBpedia	**.004**	**.004**	**.008**	**.013**	**.014**	**.024**
		Wikidata	**.008**	**.005**	.005	**.013**	**.015**	**.027**
	m.a.i.		.053		.102		.127	
MEANTIME	baseline		.777		.621		.561	
	JPARK	YAGO	**.028**	**.021**	.001	-.005	**.031**	**.025**
		DBpedia	**.030**	**.026**	-.002	.000	**.027**	**.030**
		Wikidata	**.028**	**.024**	.000	.008	**.029**	**.042**
	m.a.i.		.096		.106		.162	
TAC KBP	baseline		.760		.412		.376	
	JPARK	YAGO	**.012**	**.013**	**.007**	**.027**	**.019**	**.049**
		DBpedia	**.019**	**.015**	**.019**	**.031**	**.035**	**.054**
		Wikidata	**.016**	**.014**	**.017**	**.026**	**.034**	**.051**
	m.a.i.		.073		.172		.208	
MERGED	baseline		.838		.568		.535	
	JPARK	YAGO	**.011**	**.009**	**.006**	**.024**	**.019**	**.041**
		DBpedia	**.012**	**.010**	**.011**	**.020**	**.024**	**.037**
		Wikidata	**.013**	**.010**	**.009**	**.019**	**.024**	**.039**
	m.a.i.		.065		.133		.165	

stat. sign. results in bold

- Improvement (mostly stat. sign.) almost everywhere
- Particularly effective in selecting the correct annotation couple
- NIL strategy almost on par with NO-NIL on type, but better on link and type+link

Experiment A

CoreNLP + DBpedia Spotlight

dataset	setting	back. know.	type (F_1) =	+NIL	link (F_1) <	+NIL	type+link (F_1) <	+NIL
AIDA	baseline		.908		.656		.630	
	JPARK	YAGO	**.006**	**.005**	**.006**	**.023**	**.016**	**.035**
		DBpedia	**.004**	**.004**	**.008**	**.013**	**.014**	**.024**
		Wikidata	**.008**	**.005**	.005	**.013**	**.015**	**.027**
	m.a.i.		.053		.102		.127	
MEANTIME	baseline		.777		.621		.561	
	JPARK	YAGO	**.028**	**.021**	.001	-.005	**.031**	**.025**
		DBpedia	**.030**	**.026**	-.002	.000	**.027**	**.030**
		Wikidata	**.028**	**.024**	.000	.008	**.029**	**.042**
	m.a.i.		.096		.106		.162	
TAC KBP	baseline		.760		.412		.376	
	JPARK	YAGO	**.012**	**.013**	**.007**	**.027**	**.019**	**.049**
		DBpedia	**.019**	**.015**	**.019**	**.031**	**.035**	**.054**
		Wikidata	**.016**	**.014**	**.017**	**.026**	**.034**	**.051**
	m.a.i.		.073		.172		.208	
MERGED	baseline		.838		.568		.535	
	JPARK	YAGO	**.011**	**.009**	**.006**	**.024**	**.019**	**.041**
		DBpedia	**.012**	**.010**	**.011**	**.020**	**.024**	**.037**
		Wikidata	**.013**	**.010**	**.009**	**.019**	**.024**	**.039**
	m.a.i.		.065		.133		.165	

stat. sign. results in bold

- Improvement (mostly stat. sign.) almost everywhere
- Particularly effective in selecting the correct annotation couple
- NIL strategy almost on par with NO-NIL on type, but better on link and type+link

Experiment A

CoreNLP + DBpedia Spotlight

dataset	setting	back. know.	type (F_1)	+NIL	link (F_1)	+NIL	type+link (F_1)	+NIL
AIDA	baseline		.908		.656		.630	
	JPARK	YAGO	**.006**	.005	**.006**	.023	**.016**	.035
		DBpedia	**.004**	.004	**.008**	.013	**.014**	.024
		Wikidata	**.008**	.005	.005	.013	**.015**	.027
	m.a.i.		.053		.102		.127	
MEANTIME	baseline		.777		.621		.561	
	JPARK	YAGO	**.028**	.021	.001	-.005	**.031**	.025
		DBpedia	**.030**	.026	-.002	.000	**.027**	.030
		Wikidata	**.028**	.024	.000	.008	**.029**	.042
	m.a.i.		.096		.106		.162	
TAC KBP	baseline		.760		.412		.376	
	JPARK	YAGO	**.012**	.013	**.007**	.027	**.019**	.049
		DBpedia	**.019**	.015	**.019**	.031	**.035**	.054
		Wikidata	**.016**	.014	**.017**	.026	**.034**	.051
	m.a.i.		.073		.172		.208	
MERGED	baseline		.838		.568		.535	
	JPARK	YAGO	**.011**	.009	**.006**	.024	**.019**	.041
		DBpedia	**.012**	.010	**.011**	.020	**.024**	.037
		Wikidata	**.013**	.010	**.009**	.019	**.024**	.039
	m.a.i.		.065		.133		.165	

stat. sign. results in bold

- Improvement (mostly stat. sign.) almost everywhere
- Particularly effective in selecting the correct annotation couple
- NIL strategy almost on par with NO-NIL on type, but better on link and type+link
- Improvement consistent with all background knowledge resources

Experiment A

CoreNLP + DBpedia Spotlight

dataset	setting	back. know.	type (F_1)	+NIL	link (F_1)	+NIL	type+link (F_1)	+NIL
AIDA	baseline		.908		.656		.630	
	JPARK	YAGO	**.006**	**.005**	**.006**	**.023**	**.016**	**.035**
		DBpedia	**.004**	**.004**	**.008**	**.013**	**.014**	**.024**
		Wikidata	**.008**	**.005**	.005	**.013**	**.015**	**.027**
	m.a.i.		.053		.102		.127	
MEANTIME	baseline		.777		.621		.561	
	JPARK	YAGO	**.028**	**.021**	.001	-.005	**.031**	**.025**
		DBpedia	**.030**	**.026**	-.002	.000	**.027**	**.030**
		Wikidata	**.028**	**.024**	.000	.008	**.029**	**.042**
	m.a.i.		.096		.106		.162	
TAC KBP	baseline		.760		.412		.376	
	JPARK	YAGO	**.012**	**.013**	**.007**	**.027**	**.019**	**.049**
		DBpedia	**.019**	**.015**	**.019**	**.031**	**.035**	**.054**
		Wikidata	**.016**	**.014**	**.017**	**.026**	**.034**	**.051**
	m.a.i.		.073		.172		.208	
MERGED	baseline		.838		.568		.535	
	JPARK	YAGO	**.011**	**.009**	**.006**	**.024**	**.019**	**.041**
		DBpedia	**.012**	**.010**	**.011**	**.020**	**.024**	**.037**
		Wikidata	**.013**	**.010**	**.009**	**.019**	**.024**	**.039**
	m.a.i.		.065		.133		.165	

stat. sign. results in bold

- Improvement (mostly stat. sign.) almost everywhere
- Particularly effective in selecting the correct annotation couple
- NIL strategy almost on par with NO-NIL on type, but better on link and type+link
- Improvement consistent with all background knowledge resources
- Improvement ~10-20% of the maximum achievable one

Experiment B

flair + End-to-End Neural EL

dataset	setting	back. know.	type (F_1)	link (F_1)	type+link (F_1)
	baseline		.949	.798	.786
		YAGO	.001	**.004**	**.005**
AIDA	JPARK	DBpedia	.001	**.002**	**.003**
		Wikidata	.000	**.004**	**.004**
	m.a.i.		.033	.079	.091
	baseline		.845	.705	.669
		YAGO	.006	.001	.007
MEANTIME	JPARK	DBpedia	**.008**	.001	**.010**
		Wikidata	**.008**	.001	**.010**
	m.a.i.		.069	.024	.060
	baseline		.843	.458	.448
		YAGO	**.003**	.001	**.004**
TAC KBP	JPARK	DBpedia	**.004**	.003	**.007**
		Wikidata	**.004**	**.013**	**.017**
	m.a.i.		.033	.062	.072
	baseline		.898	.686	.672
		YAGO	**.002**	**.003**	**.005**
MERGED	JPARK	DBpedia	**.002**	**.003**	**.005**
		Wikidata	**.003**	**.006**	**.009**
	m.a.i.		.036	.069	.083

stat. sign. results in bold

- Improvement (mostly stat.sign.) in all cases except one
- Particularly effective in selecting the correct annotation couple
- Improvement consistent with all background knowledge resources
- Improvement ~10-20% of the maximum achievable one

Overall Outcome

- JPARK enables to consistently improve NERC and EL performances
  - with different combinations of NERC and EL tools
  - with different background knowledge resources
  - on three different datasets

Overall Outcome

- JPARK enables to consistently improve NERC and EL performances
  - with different combinations of NERC and EL tools
  - with different background knowledge resources
  - on three different datasets

Does the JPARK a posteriori joint revision of annotations by NERC and EL tools, performed leveraging ontological knowledge, improve NERC and EL performance?

Observations

- Very same instantiation of the model (trained on AIDA) works well on different data and with different tools
- Improvement over the baselines also on AIDA, where most of the considered tools (CoreNLP, Flair, and E2E Neural EL) were trained and developed
- Applicability to other NERC and EL tools: JPARK works on NERC and EL candidate annotations, and not on tools
  - same NERC types and linking KB: the evaluation model can be applied as-is
  - other NERC types and linking KB: reconstruct the model on an appropriate corpus G
- Implemented as a Java module of PIKES (https://pikes.fbk.eu/)
  - performance of the approach clearly depends on the size of the model
  - computation and required memory are basically negligible compared to the annotation tools

Conclusions

JPARK

- A knowledge-driven general probabilistic model for assessing and improving the coherence of NLP annotations
- Concrete implementation of the model for NERC and Entity Linking
- Comprehensive evaluation with various datasets, tools, and background knowledge resources empirically confirmed the benefits of the approach

Future Work

- NERC + EL scenario
  - Application and evaluation with other EL reference knowledge bases and NERC types (e.g., with fine-grained NERC tools)

- Application and validation of the approach to other NLP entity analyses
  - Semantic Role Labeling:

    $\boxed{\text{Kia}}$ has hired Peter Schreyer as chief design officer.

    framenet:employer

    - role annotations may imply some ontological classes (c.f., “employer” role)
  - Relation Extraction:
    - the type of the relation has implications on the subject and object entities (c.f., “born in” relation)

References

JPARK

Full results and evaluation material:
https://pikes.fbk.eu/jpark.html

Joint Posterior Revision of NLP Annotations via Ontological Knowledge
Marco Rospocher and Francesco Corcoglioniti
Proc. of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden DOI: 10.24963/ijcai.2018/600

Knowledge-driven joint posterior revision of named entity classification and linking
Marco Rospocher and Francesco Corcoglioniti
Journal of Web Semantics vol. 65:100617 (2020) DOI: 10.1016/j.websem.2020.100617

JPARK

Joint Posterior Revision of NLP Annotations via Ontological Knowledge

This page provides additional details on **JPARK**, an ontological knowledge powered probabilistic approach for jointly revising multiple NLP entity annotations.

The proposed approach is fully implemented and evaluated in the following paper:

- **Joint Posterior Revision of NLP Annotations via Ontological Knowledge**
  By Marco Rospocher and Francesco Corcoglioniti.
  In Proceedings of the 27th International Joint Conference on Artificial Intelligence and the 23rd European Conference on Artificial Intelligence, IJCAI-ECAI 2018, Stockholm, Sweden, July 13-19, 2018
  [bib] [pre-print/mirror]

JPARK has been evaluated on three reference datasets for Named Entity Recognition and Classification (NERC) and Entity Linking (EL):

- AIDA CoNLL-YAGO: This dataset consists of 1,393 English news wire articles from Reuters, with 34,999 mentions hand-annotated with named entity types (PER, ORG, LOC, MISC) for the CONLL2003 shared task on named entity recognition, and later hand-annotated with the YAGO2 entities and corresponding Wikipedia page URLs. It is split in three parts: eng.train (946 docs), eng.testa (216 docs), eng.testb (231 docs).

Joint Posterior Revision of NLP Annotations via Ontological Knowledge

Marco Rospocher, Francesco Corcoglioniti

IJCAI ECAI 2018

Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence
Main track. Pages 4316-4322. https://doi.org/10.24963/ijcai.2018/600

PDF BibTeX

Different well-established NLP tasks contribute to elicit the semantics of entities mentioned in natural language text, such as Named Entity Recognition and Classification (NERC) and Entity Linking (EL). However, combining the outcomes of these tasks may result in NLP annotations --- such as a NERC organization linked by EL to a person --- that are unlikely or contradictory when interpreted in the light of common world knowledge about the entities these annotations refer to. We thus propose a general probabilistic model that explicitly captures the relations between multiple NLP annotations for an entity mention, the ontological entity classes implied by those annotations, and the background ontological knowledge those classes may be consistent with. We use the model to estimate the posterior probability of NLP annotations given their confidences (prior probabilities) and the ontological knowledge, and consequently revise the best annotation choice performed by the NLP tools. In a concrete scenario with two state-of-the-art tools for NERC and EL, we experimentally show on three reference datasets that for these tasks, the joint annotation revision performed by the model consistently improves on the original results of the tools.

Keywords:
Natural Language Processing: NLP Applications and Tools
Natural Language Processing: Knowledge Extraction
Natural Language Processing: Named Entities

Web Semantics: Science, Services and Agents on the World Wide Web 65 (2020) 100617

Contents lists available at ScienceDirect

Web Semantics: Science, Services and Agents on the World Wide Web

journal homepage: www.elsevier.com/locate/websem

Knowledge-driven joint posterior revision of named entity classification and linking

Marco Rospocher [a,*], Francesco Corcoglioniti [b]

[a] *Università degli studi di Verona, Lungadige Porta Vittoria, 41 - 37129 Verona, Italy*
[b] *Free University of Bozen-Bolzano, Piazza Università, 1 - 39100 Bolzano, Italy*

ARTICLE INFO

Article history:
Received 25 September 2019
Received in revised form 24 August 2020
Accepted 6 October 2020
Available online 10 October 2020

ABSTRACT

In this work we address the problem of extracting quality entity knowledge from natural language text, an important task for the automatic construction of knowledge graphs from unstructured content.

More in details, we investigate the benefit of performing a joint posterior revision, driven by ontological background knowledge, of the annotations resulting from natural language processing (NLP) entity analyses such as named entity recognition and classification (NERC) and entity linking (EL). The revision is performed via a probabilistic model, called JPARK, that given the candidate annotations independently identified by NERC and EL tools on the same textual entity mention, reconsiders the best annotation choice performed by the tools in light of the coherence of the candidate annotations with the ontological knowledge. The model can be explicitly instructed to handle the information that an entity can potentially be NIL (i.e., lacking a corresponding referent in the target linking knowledge base), exploiting it for predicting the best NERC and EL annotation combination.

We present a comprehensive evaluation of JPARK along various dimensions, comparing its performances with and without exploiting NIL information, as well as the usage of three different background knowledge resources (YAGO, DBpedia, and Wikidata) to build the model. The evaluation, conducted using different tools (the popular Stanford NER and DBpedia Spotlight, as well as the more recent Flair NER and End-to-End Neural EL) with three reference datasets (AIDA, MEANTIME, and TAC-KBP), empirically confirms the capability of the model to improve the quality of the annotations of the given tools, and thus their performances on the tasks they are designed for.

JPARK
pikes.fbk.eu/jpark

KE4IR
pikes.fbk.eu/ke4ir

PIKES
pikes.fbk.eu

PSL4EA
pikes.fbk.eu/psl4ea

BPMN Ontology
dkm.fbk.eu/bpmn-ontology

PREMON
premon.fbk.eu

PESCaDO
bit.ly/pescado-onto

NewsReader
Event & Situation Ontology
github.com/newsreader/eso

Marco Rospocher

http://marcorospocher.com/ marco.rospocher@univr.it @marcorospocher

UNIVERSITÀ di VERONA
Dipartimento di LINGUE E LETTERATURE STRANIERE
Dipartimento di Eccellenza 2018/2022 MIUR

Slides

RDFpro
rdfpro.fbk.eu

KnowledgeStore
knowledgestore.fbk.eu

TeX-Owl
github.com/dkmfbk/TexOwl

MoKi
the Modelling WiKi ---
moki.fbk.eu