

Overview of TAC-KBP2020 Recognizing Ultra Fine-grained Entities Task (RUFES)

Heng Ji (UIUC), Avirup Sil (IBM Research), Shudong Huang (NIST),
Hoa Trang Dang (NIST), Ian Soboroff (NIST)

NIST



IBM Research

Motivations and Goals

- **Questions for a knowledge base often involve fine-grained knowledge elements, e.g.,**
 - “Which epidemiologists have attended the meeting?”
 - “Which **amino acids** in glycoprotein are most related to Glycan?”
- **Limitations of Existing fine-grained entity extraction benchmarks**
 - The amount of gold-standard annotation is very low; while most annotations are ‘silver-standard’, derived from automatic mapping from KBs onto unstructured texts
 - Most data sets are Wikipedia articles instead of other documents in the real-world that may contain unknown entities
 - Only assign one type per entity mention given a certain context
 - The existing ontologies are not designed based on the distributions of entities in data from a wide range of topics
 - Existing methods have focused on name mentions only.
 - There are very few shared open-source systems
- **Our goals**
 - Significantly extend entity extraction capabilities to a wider variety of fine-grained entity types (e.g., technical terms, lawsuits, disease, crisis, vehicles, food, biomedical entities)
 - Perform document-level knowledge aggregation

Task definition 2020

- A new task!
 - RUFES (pronounced *Roofes*): Recognizing Ultra-fine grained Entities
- What's new in RUFES?
 - New source data: Washington Post collection (2012-2019)
 - Reorganized and expanded ontology
 - Added nominal and pronominal mentions
 - Paragraph as context for entity types
 - Allow multiple entity types
 - Document-level coreference resolution

Task Overview

- Oct . 18 in Bridgeview : Costa Rica vs . Jamaica , 4 ; **Martinique** vs . Mexico , 6 : 30

EntityID: 1
Type: ORG.Association.SportsTeam

- **Martinique** is not a full member of FIFA and not eligible to qualify for the World Cup . (The qualifying tournament is , in fact , the CONCACAF championship , for which Martinique qualified through Caribbean competition .) If Martinique advances to the semifinals , the fifth - place team will remain in contention for a World Cup berth .

- In recent weeks , another mosquito - borne virus common in Africa and Asia has spread quickly through the eastern Caribbean , appearing for the first time in the Western Hemisphere . Chikun gunya fever , which is similar to dengue , was reported in December on the French side of St . Martin and has spread to seven other jurisdictions , including **Martinique** , Guadeloupe and the British Virgin Islands .

EntityID: 2
Type: GPE.Territory

Wider context info is necessary in RUFES

- Last year , the CDC reported only seven confirmed norovirus outbreaks , which came to just 1 , 238 total afflicted passengers worldwide , or 0 . 0059 percent of 21 million cruise **passengers** . In 2012 , the agency reported 16 norovirus outbreaks on ships .

EntityID: 3
Type: PER.SituationalRole.Patient

Finer context modeling is necessary in RUFES

Human Feedback

- Simulate how analysts might interact with the system:
- The user is reviewing a document and working through it from the beginning to the end. During this sequential process, if the user encounters an incorrect system-generated annotation (which could be a miss, an error in mention extent, a type error, or a coreference error), the user tags the error and provides the correction
- The user has some tolerance value k for the amount of error they are willing to see before they lose confidence in the system's annotations and stop reviewing the document
- All participants in EDL Evaluation Window 1 were given feedback on the same set of 10 documents
 - Feedback was given until $k=40$ errors were detected in the document
- The 10 documents for feedback were selected to be documents where the runs submitted to Evaluation Window 1 had low confidence

Datasets

	Development	Sample Annotation	Evaluation	Evaluation Annotation
Planned	100,000	50	100,000	250
Delivered	100,000	50	100,000	106

- Long articles are truncated so that the average length (based on the annotated data) is 500 words
- 139 first-passed from the evaluation set
- 106 completed second pass when the second evaluation window closed
- 33 remaining articles reserved for 2021

Ontology Development

- Goal: over 250 entity types (coarse, fine and ultra-fine combined)
- AIDA ontology as base (179)
 - Some dropped; some reorganized
 - SituationalRole*
- Expansion with input from YAGO and Wikidata
 - Scan news articles for potential entity types salient to the topic
 - New top-level class: APP, Pathogen, Publication

Entities: AIDA vs RUFES vs Annotated

	Level 1	Level 2	Level 3	Total	Level 1 2 Terminal
AIDA	17	61	101	179	24
RUFES	14	80	172	266	37
Annotated	14	79	146	239	

Mention Types and Reference

- Added nominals and pronominals
- Removed specific reference limitation
- Mentions are taggable if they are referential
 - Singular specific reference
 - Plural specific (group) reference
 - Generic
 - Indefinite
- Limitations of tagging named mentions only:
 - Named mentions usually refer to specific individual entities
 - Many entities do not usually have names at the individual level, although named mentions may make “class/type/kind” reference to those entities

Annotated Mention Types

Mention Level

	NAM	NOM	PRO	Total
Mention	5869	6270	2621	14760

Entity Level

	Total	NAM	NOM	PRO	NOM w/o NAM	PRO Only
Entity	6299	2732	4145	1257	3333	234

Participants (alphabetical order)

- 24 teams registered; 5 teams participated

Team	Affiliation
ibm_mnlp_ie	IBM Research (US)
L3i	La Rochelle Université (France)
RelMatter	Northwestern Polytechnical University (China)
sinai	Universidad de Jaén (Spain)
UIUC	University of Illinois at Urbana-Champaign (US)

Performance – Mention Detection

Team	strong_mention_match			strong_typed_mention_match*		
	P	R	F	P	R	F
L3i	0.869	0.867	0.868	0.743	0.740	0.741
ibm_mnlp_ie	0.829	0.848	0.839	0.715	0.729	0.722
UIUC	0.774	0.803	0.789	0.679	0.695	0.687
RelMatter	0.825	0.786	0.805	0.345	0.335	0.340
sinai	0.388	0.477	0.428	0.172	0.210	0.189

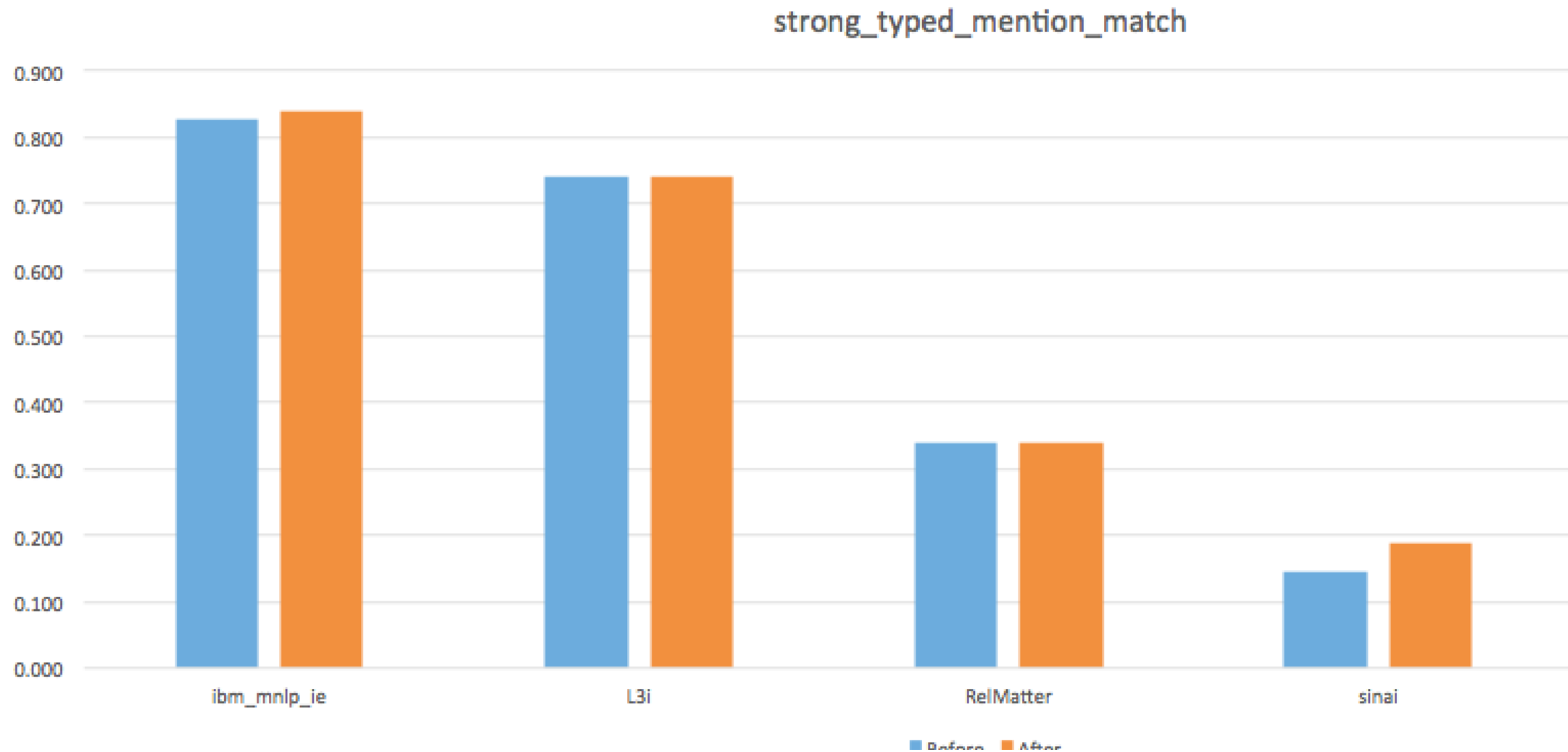
Performance – Coreference

Team	mention_ceaf			typed_mention_ceaf*			type metric**
	P	R	F	P	R	F	F
UIUC	0.677	0.702	0.689	0.614	0.631	0.622	0.383
ibm_mnlp_ie	0.584	0.598	0.591	0.536	0.548	0.542	0.417
L3i	0.579	0.577	0.578	0.504	0.503	0.504	0.323
RelMatter	0.522	0.497	0.509	0.226	0.215	0.221	0.234
sinai	0.238	0.292	0.262	0.133	0.164	0.147	0.061

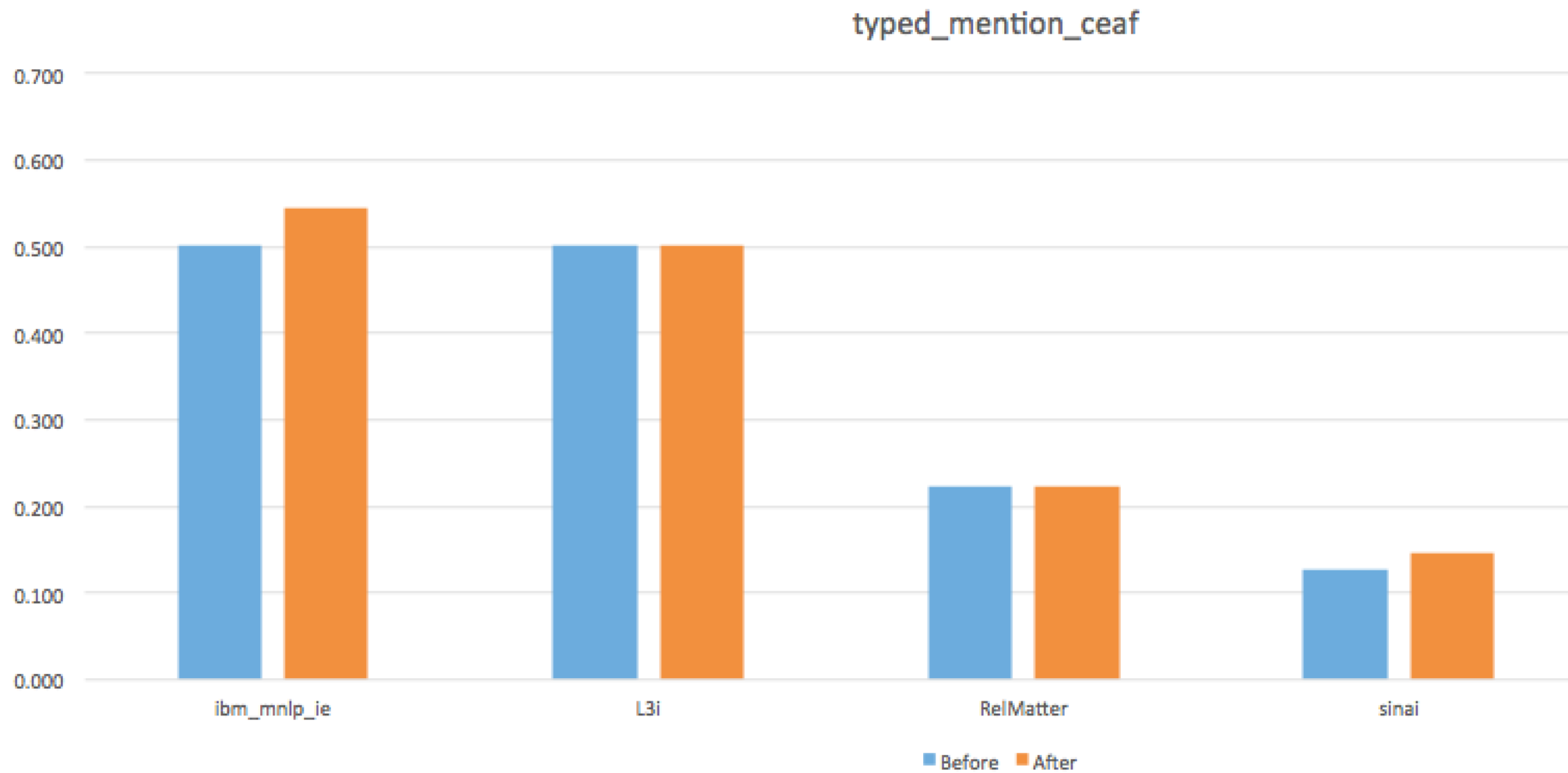
*Top-level entity types

**New fine-grained entity type metric

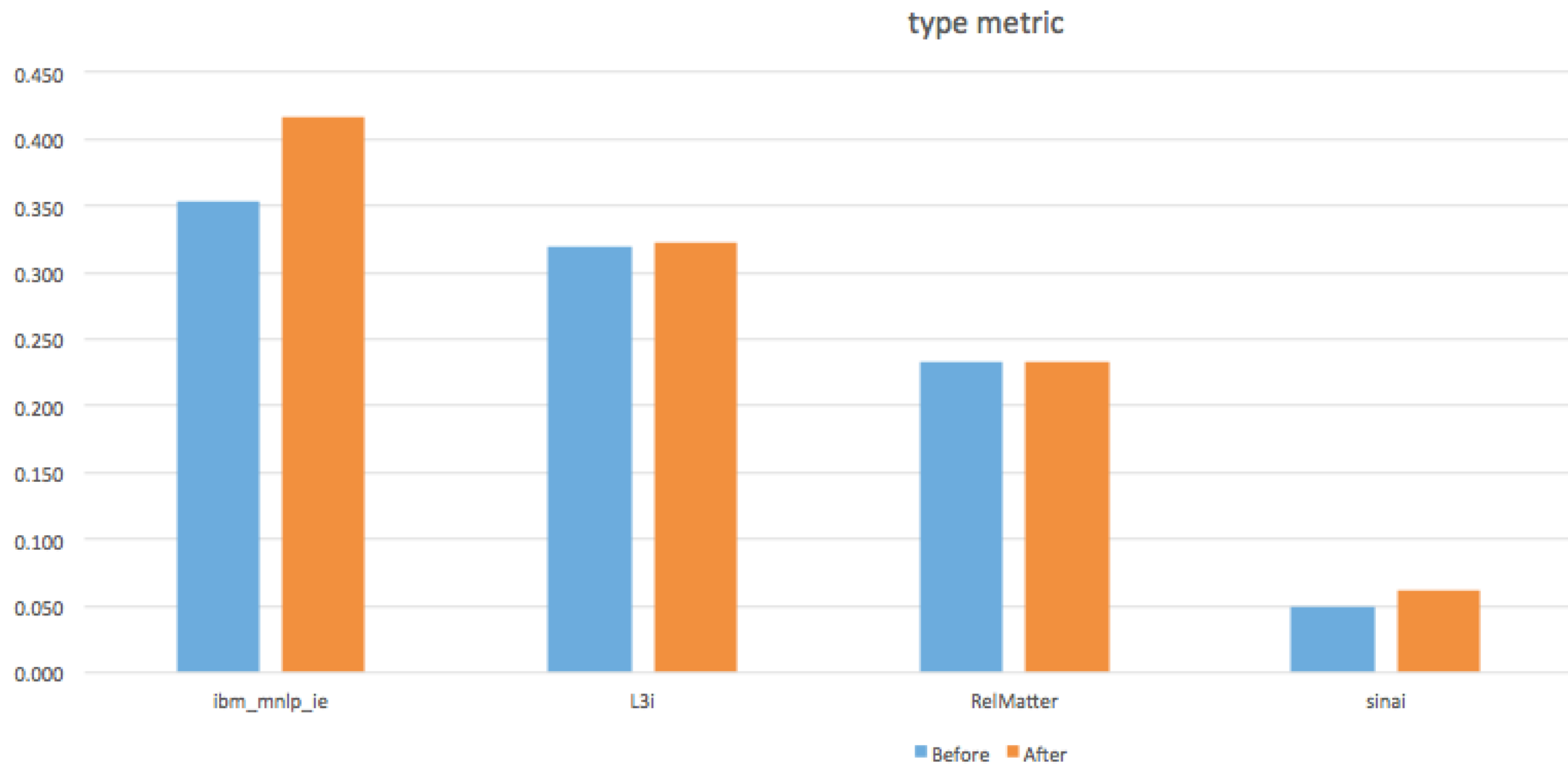
Impact of Human Feedback (Mention)



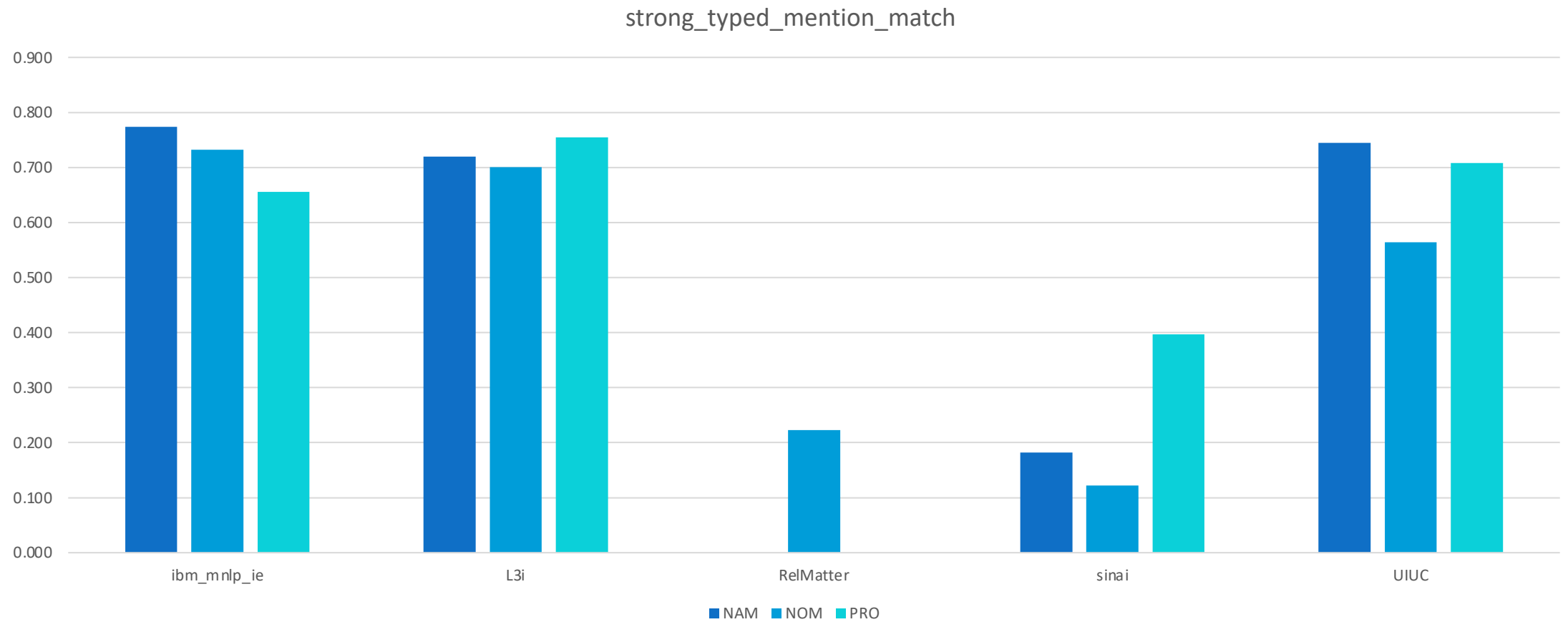
Impact of Human Feedback (Coreference)



Impact of Human Feedback (Fine-Grain Typing)



Mention Types

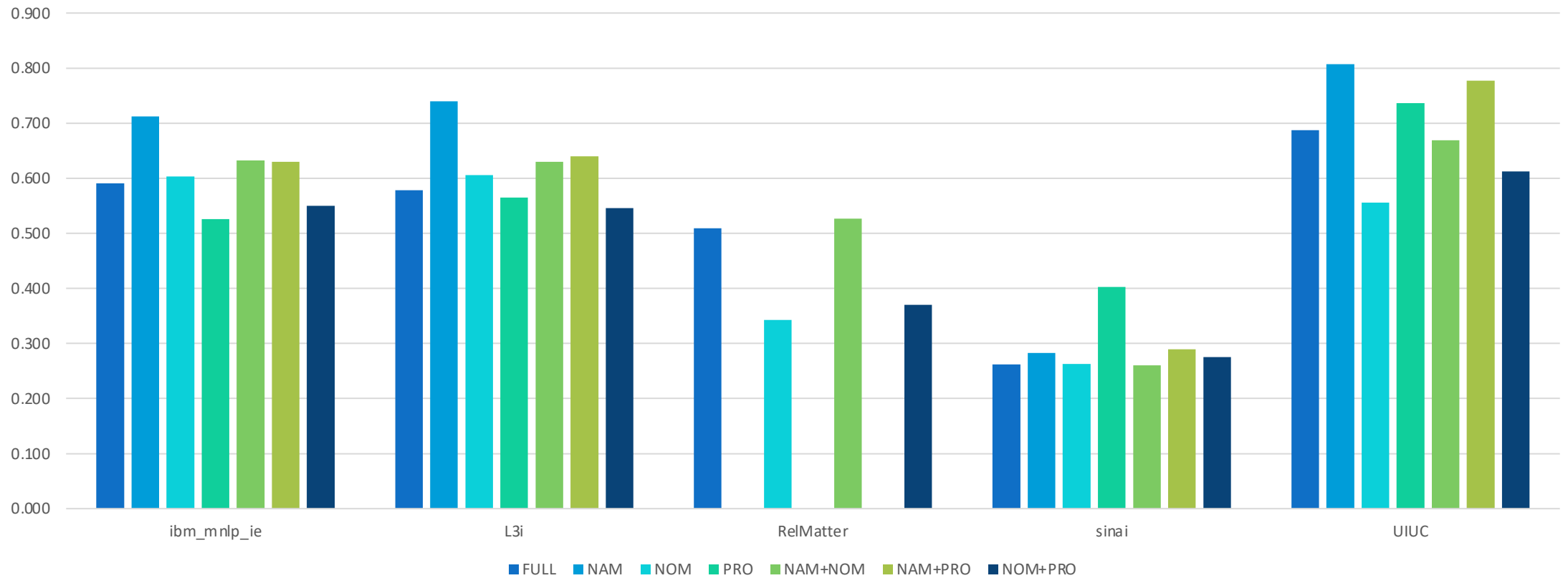


Mention Types and Coreference

Team	FULL	NAM	NOM	PRO	NAM+NOM	NAM+PRO	NOM+PRO
UIUC	0.687	0.807	0.556	0.736	0.669	0.777	0.612
ibm_mnlp_ie	0.591	0.712	0.603	0.526	0.632	0.630	0.550
L3i	0.578	0.740	0.606	0.565	0.630	0.640	0.546
RelMatter	0.509	0.000	0.343	0.000	0.527	0.000	0.370
sinai	0.262	0.283	0.263	0.403	0.260	0.289	0.275

Mention Types and Coreference

mention_ceaf

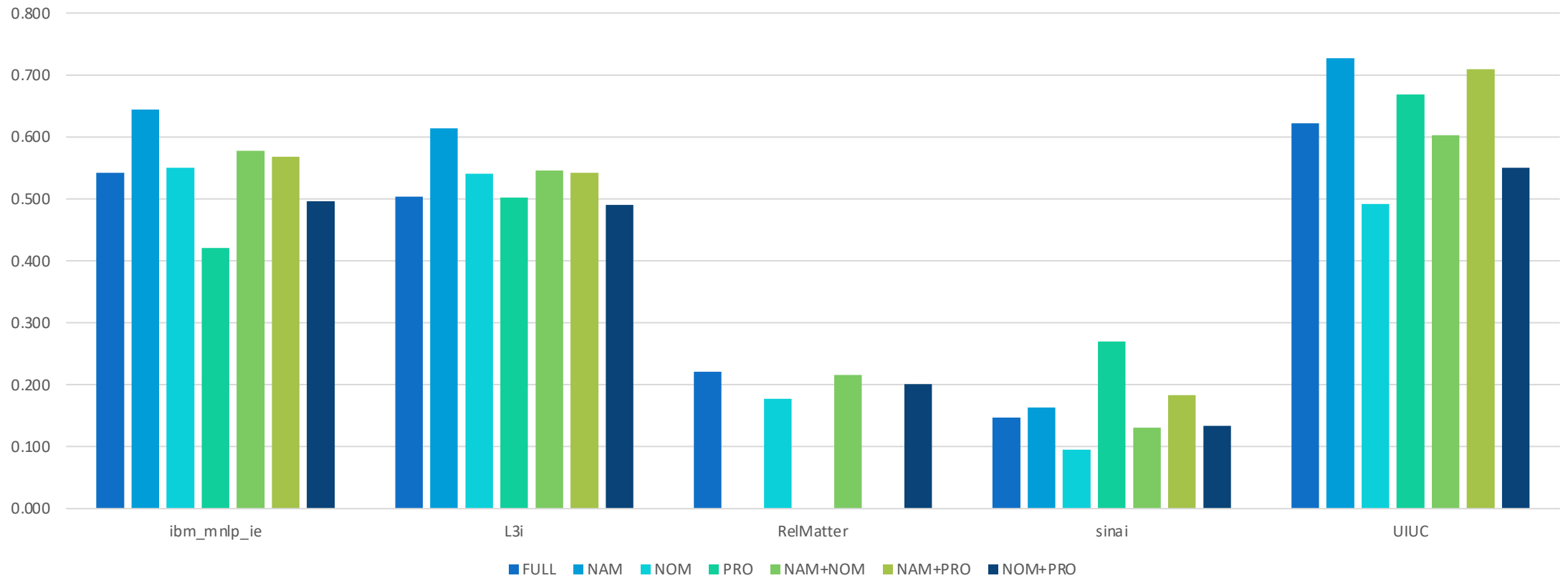


Mention Types, Entity Types and Coreference

Team	FULL	NAM	NOM	PRO	NAM+NOM	NAM+PRO	NOM+PRO
UIUC	0.622	0.727	0.492	0.669	0.603	0.709	0.550
ibm_mnlp_ie	0.542	0.644	0.550	0.421	0.578	0.568	0.496
L3i	0.504	0.614	0.541	0.502	0.546	0.542	0.490
RelMatter	0.221	0.000	0.177	0.000	0.216	0.000	0.201
sinai	0.147	0.163	0.095	0.270	0.131	0.183	0.134

Mention Types, Entity Types and Coreference

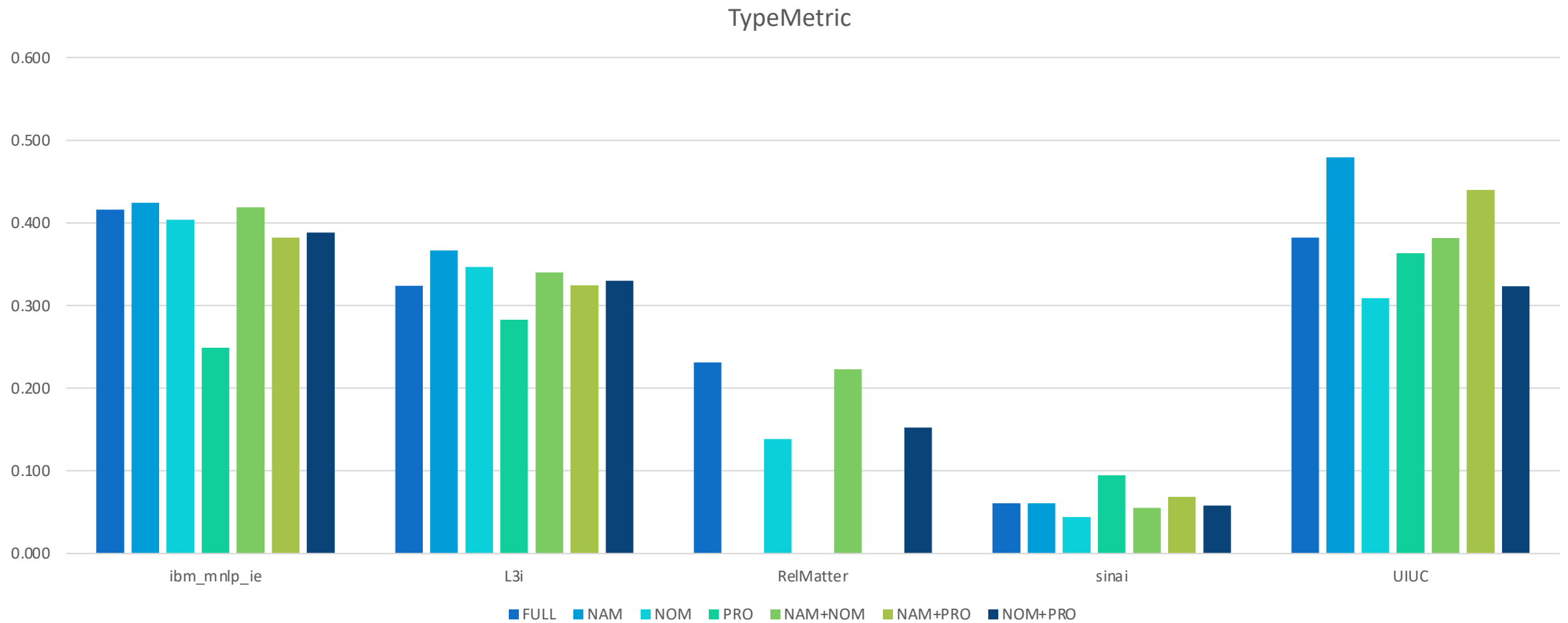
typed_mention_ceaf



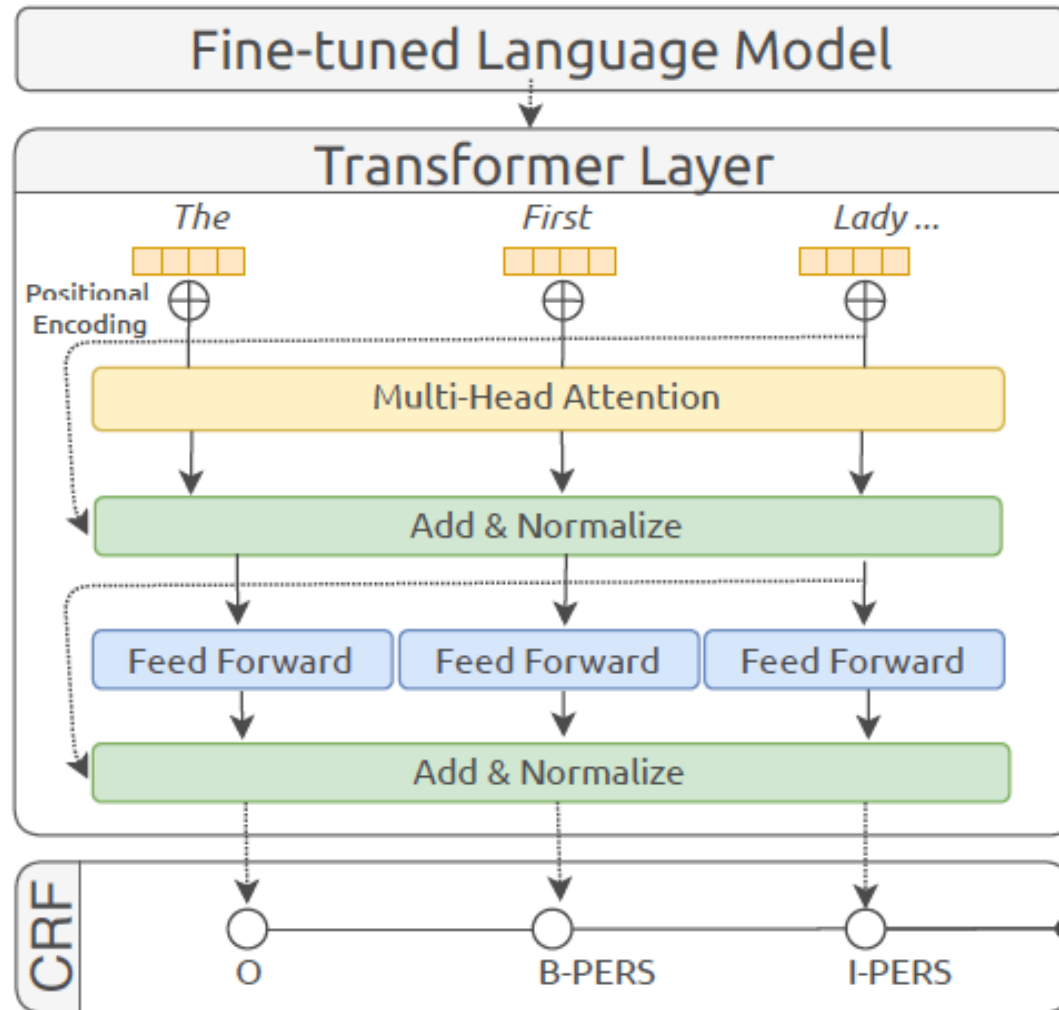
Mention Types and Fine Typing

Team	FULL	NAM	NOM	PRO	NAM+NOM	NAM+PRO	NOM+PRO
ibm_mnlp_ie	0.416	0.424	0.404	0.249	0.419	0.382	0.388
UIUC	0.382	0.479	0.309	0.364	0.382	0.440	0.324
L3i	0.324	0.367	0.347	0.283	0.340	0.324	0.330
RelMatter	0.231	0.000	0.139	0.000	0.223	0.000	0.153
sinai	0.061	0.061	0.044	0.095	0.056	0.069	0.058

Mention Types and Fine Typing



Successful Methods: L3i Hierarchical Multitask Learning for Entity Extraction

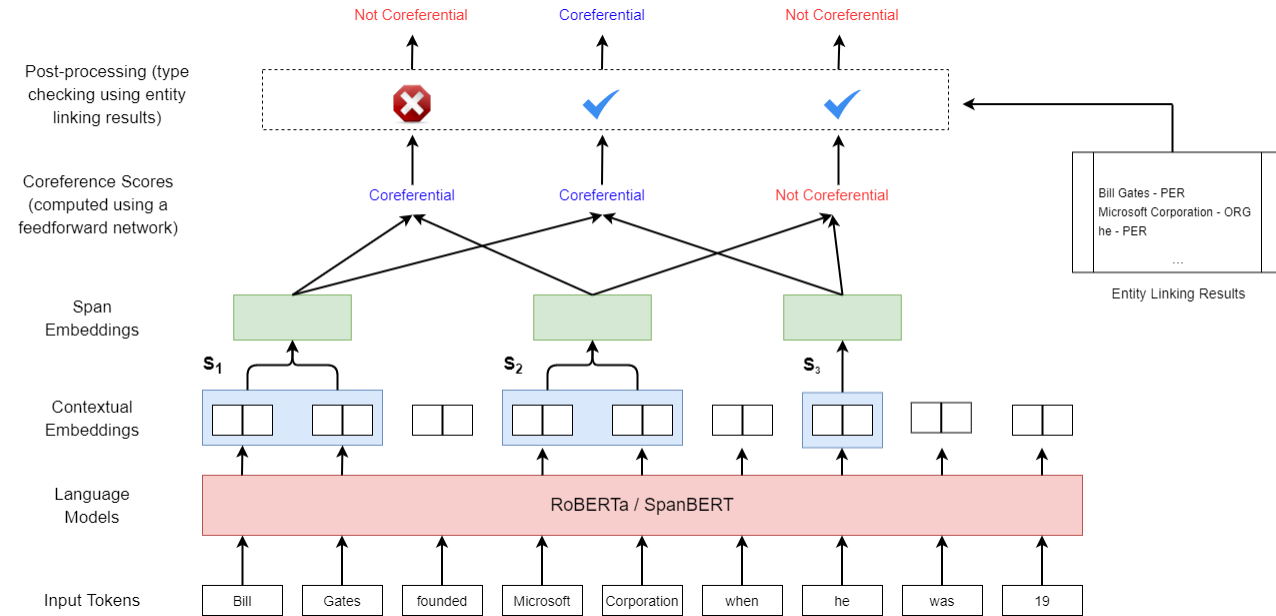


Successful Methods: IBM Synthetic Data Augmentation

- Use IBM in-house KLUE dataset (Florian et al., 2004) and DBPedia Abstract Corpus to train the seed supervised model
- Generate synthetic data from the unannotated development corpus to augment the available data with k-Nearest Neighbor Graph
- Query the K-NNG with each extracted mention representation and retrieve the k nearest mentions and labels associated with them, then select the fine-grained type using a threshold to vote between the first supervised model and the second model with data augmentation

Successful Methods: UIUC Joint Entity and Event Coreference Resolution

- Adopt a span-based architecture similar to previous work ([Lee et al., 2017](#))
 - **Contextual Embeddings** ~ Using a Transformer encoder
 - **Span Representations** ~ Using an attention mechanism
 - **Pairwise Scores** ~ Using a 2-layer multilayer feedforward network
- Refine prediction decisions by using entity linking results



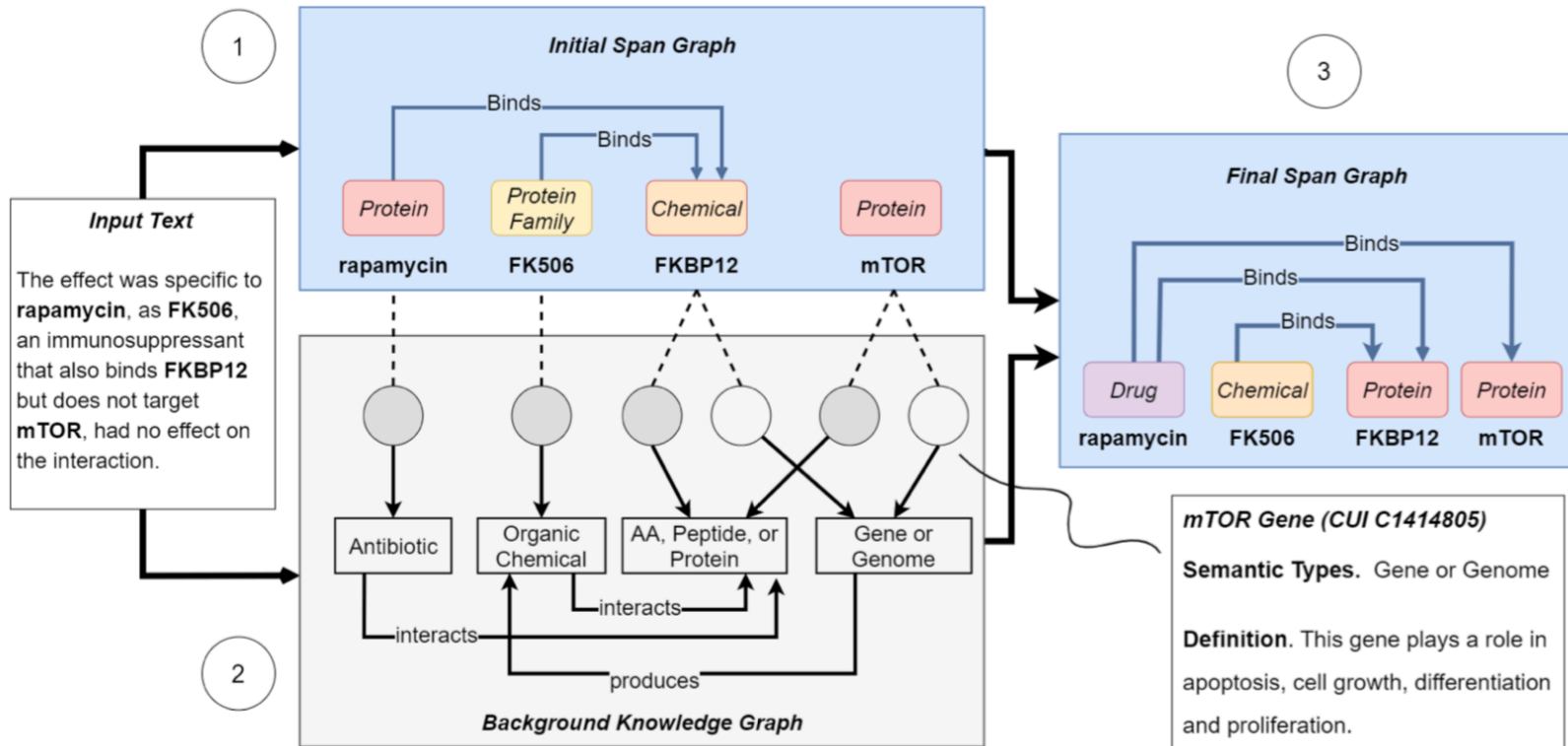
	Ontonotes (CoNLL)
Previous SOTA (Lu and Ng, 2020)	91.9
Ours	92.4

Performance of English systems (on gold mentions)

	SemEval 2010
TANL-1 (Attardi et al., 2010)	56.1
Ours	67.6

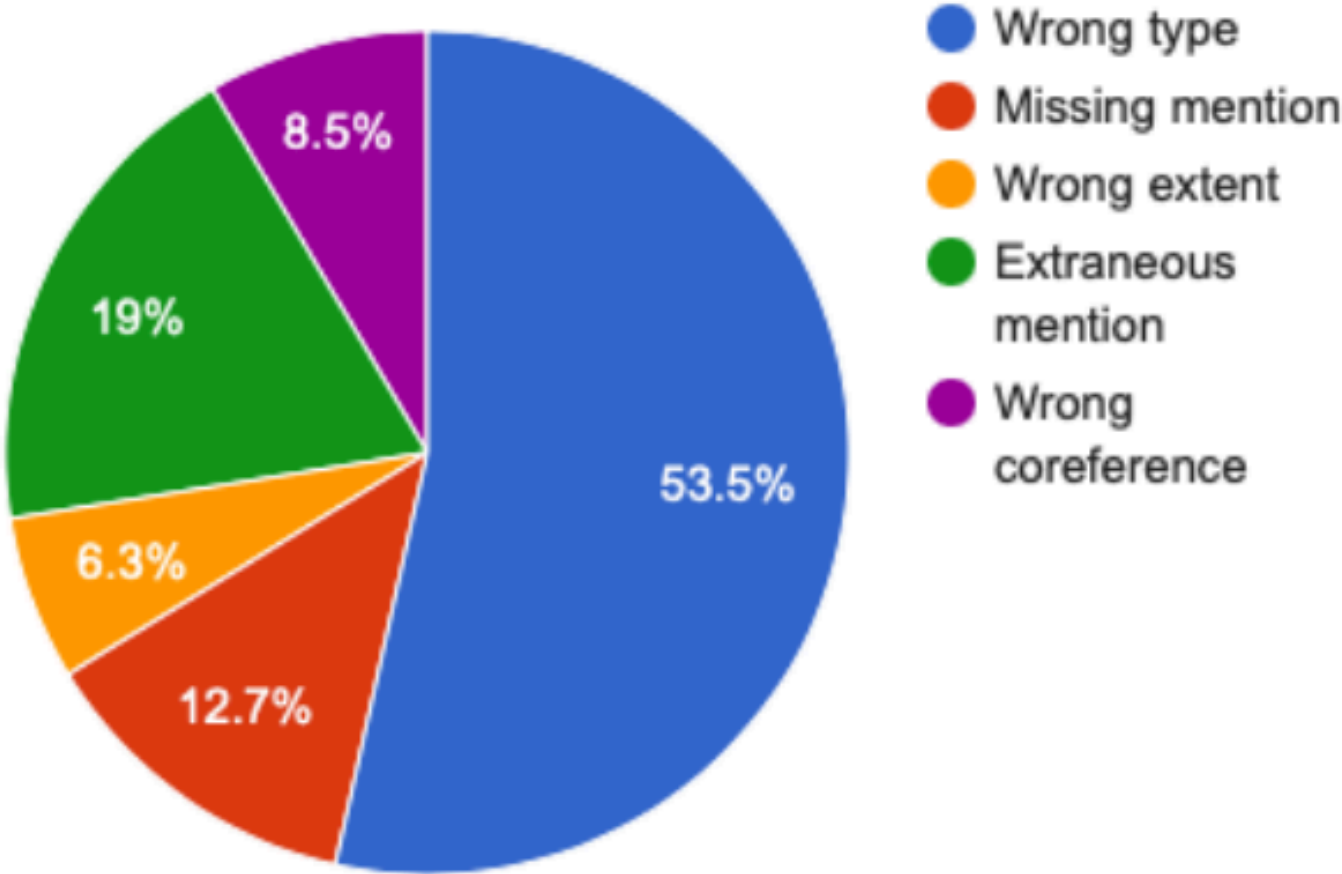
Performance of Spanish systems (on gold mentions)

Successful Methods: UIUC External Knowledge Acquisition for Entity Extraction by Entity Linking



Method	Entity (Micro-F1 %)	Relation (Micro-F1%)
Full Model (KECI)	87.42	66.09
• w/o external knowledge	83.98	63.90
• w/o collective inference	84.32	64.23

Remaining Errors: UIUC System Error Distribution



Remaining Challenges: External Knowledge

- Requires coreference resolution before typing
 - ORG.WorldOrganization instead of ORG.Government.Agency:
 - The WHO , [the U. N. Food and Agriculture Organization](#), and the World Organization for Animal Health are also participating in the effort .
- Requires entity linking before typing
 - GPE.Territory instead of GPE.country
 - In recent weeks , another mosquito - borne virus common in Africa and Asia has spread quickly through the eastern Caribbean , appearing for the first time in the Western Hemisphere . Chikungunya fever , which is similar to dengue , was reported in December on the French side of St . Martin and has spread to seven other jurisdictions , including Martinique , [Guadeloupe](#) and the British Virgin Islands .

Remaining Challenges: Better Coreference Resolution

[MENTION:] States in the upper Midwest and the Northeast will receive the bulk of the funding . New York will get an additional \$ 50 million , while Illinois , Michigan , Ohio and Pennsylvania will each receive more than \$ 20 million in new funds . Every [state](#) in the Union except Hawaii will receive at least seven figures in new funding ; Hawaii stands to gain an additional \$ 680 , 874 in LIHEAP funding , HHS said .

[GOLD:] [GPE.ProvinceState](#)

[MENTION:] “ Ambassador - designate Stanton is fully briefed on the case ,” the official said . “ Having an Ambassador in place will help elevate the U . S . [government](#) ’ s engagement with Timor - Leste authorities to ensure treatment consistent with internationally recognized standards and to ensure that the case proceeds expeditiously under local laws and international standards .”

[GOLD:] [ORG.Government](#)

[MENTION:] That move would likely be opposed by Washington , [which](#) is still attempting to “ deconflict ” its operations in Syria from those of the Russians .

[GOLD:] [ORG.Government](#)

[MENTION:] Oct . 20 : in Washington : Trinidad & Tobago vs . Guatemala , 5 ; [United States](#) vs . Haiti , 7 : 30

[GOLD:] [ORG.Association.SportsTeam](#)

Remaining Challenges: Incorporating Relation Extraction

- PER.Serviceman.MilitaryOfficer
 - According to Lt . Col . [Peter Lerner](#) , senior spokesman for the Israeli Defense Forces , a Palestinian man in his 20s approached a checkpoint at Tapuah Junction south of Nablus in the West Bank about 10 : 30 a . m . Friday .
- PER.Politician.Legislator
 - D . C . Council member [Tommy Wells](#) (D - Ward 6), who introduced the resolution to request a congressional hearing on the project , said the reconstruction of the tunnel raises concerns about rail safety and the passage of hazardous materials through the city . He said it provides an opportunity to reexamine alternatives for rerouting cargo trains from he city .

Remaining Challenges: Incorporating Event Argument Role Labeling

- PER.SituationalRole.Victim
 - There was an uptick last year in attacks against [Israelis](#) by Palestinians , who acted on their own out of personal vendettas or frustrations.
- PER.SituationalRole.Survivor
 - The last suicide bombings in Israel that led to deaths or injuries occurred in 2008 . Since then there have been numerous uses of improvised explosive devices , left along the Israeli fence that surrounds the Gaza Strip and other locales , designed to kill or maim Israeli military patrols . In 2012 , a bomb left on a Tel Aviv bus that was remotely detonated injured two dozen [people](#).
- ORG.AdvocacyGroup
 - National Train Day activities will keep Union Station hopping , with model - train displays and the Chuggington Kids Depot , where kids can do train - related crafts and play with toy locomotives . Dovetail Games will offer free downloads of the Amtrak Pacific Surfliner version of the video game “ Train Simulator 2014 ,” and people from [Operation Lifesaver](#) will be on hand to talk about how to stay safe around railroad tracks .

Remaining Challenges: Require Situation Understanding beyond sentence-level

- PER.Terrorist
 - The border guards ordered the man to stop , open his jacket and lift up his shirt .
 - He revealed “ 12 steel pipes connected with wires ” around his waist , Lerner said .
 - Later , a bomb disposal team detonated one of the metal tubes and confirmed that it contained explosive material .
- WEA.ExplosiveDevice.Bomb
 - Israeli media quoted unnamed military sources as saying that it was unclear whether the man with the explosives was planning to detonate the device at the checkpoint or travel to another location .
 - The last suicide bombings in Israel that led to deaths or injuries occurred in 2008 . Since then there have been numerous uses of improvised explosive devices , left along the Israeli fence that surrounds the Gaza Strip and other locales , designed to kill or maim Israeli military patrols . In 2012 , a bomb left on a Tel Aviv bus that was remotely detonated injured two dozen people .
- VEH.WheeledVehicle.RailedVehicle
 - From the [Hogwarts Express](#) to “ The Polar Express ,” trains transport us to magical places in books and movies . But real - life trains can be pretty magical , too . And this Saturday is all about trains . National Train Day , celebrated at hundreds of locations across the United States — including Washington ’ s Union Station — has tons of fun stuff for train fanatics and choo - choo newbies alike .

Remaining Challenges: Confusion between types

- LOC.GeographicPoint.Address instead of FAC.Way
- According to Lt . Col . Peter Lerner , senior spokesman for the Israeli Defense Forces , a Palestinian man in his 20s approached a checkpoint at [Tapuah Junction](#) south of Nablus in the West Bank about 10 : 30 a . m . Friday .
- WEA.ExplosiveDevice instead of WEA.ExplosiveDevice.Bomb
- JERUSALEM — Israeli security officials said their border police at a military checkpoint in the West Bank detained a young Palestinian man wearing a “ suicide belt ” packed with [explosives](#) . The man was arrested and did not activate the device , they said .

Plans for KBP2021

- Leaderboard: Put RUFES2020 up for open window evaluation
 - Feedback task on leaderboard may require submission of dockers that can automatically ingest feedback to produce post-feedback run (limited number of feedback runs allowed).
- Build Dockerhub for participants to share dockers
- Introduce streaming data?
 - Streaming feedback at set timepoints in the stream?
- Add domain adaptation (zero-shot transfer learning) as a focused challenge?
- Focus on improving person role extraction?
- Add a foreign language Chinese?

Thanks to TAC-KBP Scientific Board

- Mausam (Indian Institute of Technology Delhi)
- Isabelle Augenstein (University of Copenhagen)
- Elizabeth Boschee (Information Sciences Institute)
- Laura Dietz (University of New Hampshire)
- Radu Florian (IBM Research AI)
- Alan J. Goldschen (U.S. Department of Defense)
- Ralph Grishman (New York University)
- Hanna Hajishirzi (University of Washington)
- Ed Hovy (U.S. Department of Defense)
- Yunyao Li (IBM Research AI)
- Andrew McCallum (University of Massachusetts Amherst)
- Paul McNamee (Johns Hopkins University)
- Graham Neubig (Carnegie Mellon University)
- Boyan Onyshkevych (U.S. Department of Defense)
- Marius Pasca (Google)
- Siddharth Patwardhan (Apple)
- Dan Roth (University of Pennsylvania)
- Xiang Ren (University of Southern California)
- Satoshi Sekine (RIKEN Center for Advanced Intelligence)
- Sameer Singh (University of California Irvine)