Overview of TAC-KBP2017 13 Languages Entity Discovery and Linking

Heng Ji, Xiaoman Pan, Boliang Zhang, Joel Nothman, James Mayfield, Paul McNamee and Cash Costello

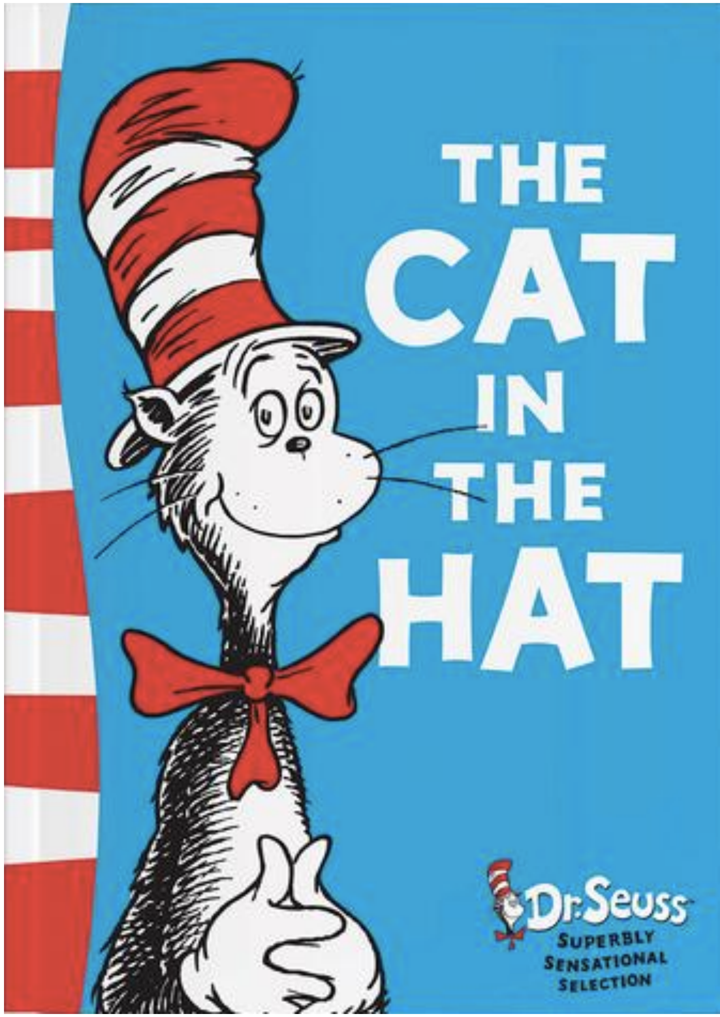
jih@rpi.edu

Thanks to KBP2016 Organizing Committee

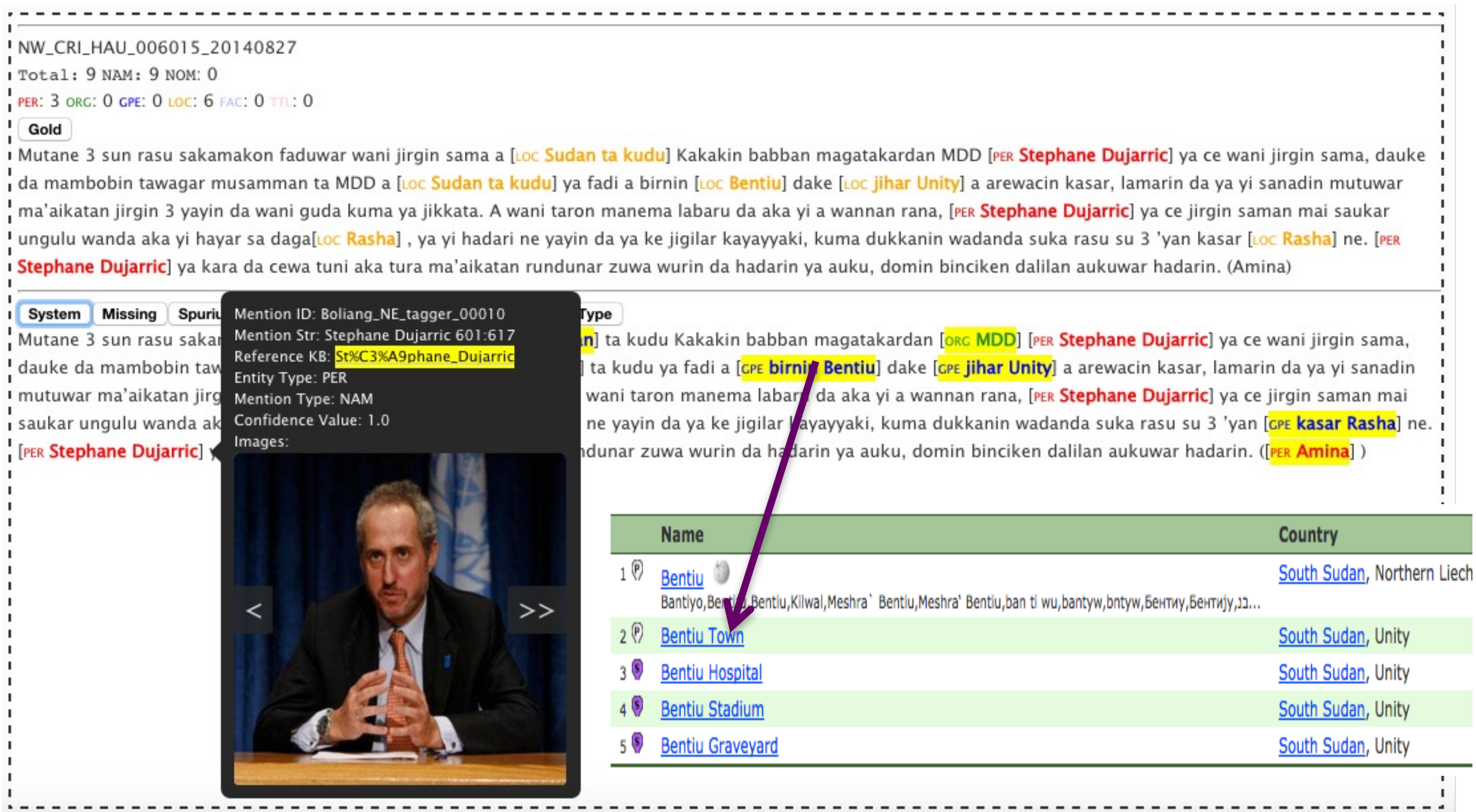
Overview Paper: http://nlp.cs.rpi.edu/kbp2017.pdf

Rensselaer

Goals and The Task



Cross-lingual Entity Discovery and Linking



Where are We Now: Awesome as Usual

- Great participation (24 teams)
- Improved Quality
  - Almost perfect linking accuracy for linkable mentions (?)
  - Almost perfect NIL clustering (?)
  - Chinese EDL 4% better than English EDL
- Improved Portability
  - 5 types of entities → 16,000 types
  - 1-3 languages → 3,000 languages
  - Scarce KBs (Geoname, World Factbook, Name List)
- Improved Scalability
  - 90,000 documents

The Tasks

- Input
  - A set of multi-lingual text documents (main task: English, Chinese and Spanish)
- Output
  - Document ID, mention ID, head, offsets
  - Entity type: GPE, ORG, PER, LOC, FAC
  - Mention type: name, nominal
  - Reference KB link entity ID, or NIL cluster ID
  - Confidence value
- A new pilot study on 10 low-resource languages
  - Polish, Chechen, Albanian, Swahili, Kannada, Yoruba, Northern Sotho, Nepali, Kikuyu and Somali
  - No NIL clustering
  - No FAC
  - No Nominal
  - KB: 03/05/16 Wikipedia dump instead of BaseKB

Evaluation Measures

Short name	Name in scoring software	Filter	Key	Evaluates
Mention evaluation				
NER	strong_mention_match	NA	*span*	Identification
NERC	strong_typed_mention_match	NA	*span,type*	+ classification
Linking evaluation				
NERLC	strong_typed_all_match	NA	*span,type,kbid*	+ linking
NELC	strong_typed_link_match	*is linked*	*span,type,kbid*	Link recognition and classification
NENC	strong_typed_nil_match	*is nil*	*span,type*	NIL recognition and classifciation
Tagging evaluation				
KBIDs	entity_match	*is linked*	*docid,kbid*	Document tagging
Clustering evaluation				
CEAFm	mention_ceaf	NA	*span*	Identification and clustering
CEAFmC	typed_mention_ceaf	NA	*span,type*	+ classification
CEAFmC+	typed_mention_ceaf_plus	NA	*span,type,kbid*	+ linking

- CEAFmC+: end to end metric for extraction, linking and clustering

Data Annotation and Resources

- Tr-lingual EDL details in LDC talk and resource overview paper (Getman et al., 2017)
- 10 Languages Pilot (Silver-standard+ prepared by RPI and JHU Chinese Rooms, adjudicated annotations by five annotators)

Languages	Training	Test	Data Source
Albanian	40 documents	10 documents	Silver+
Chechen	83 documents	30 documents	Gold
Kannada	40 documents	10 documents	Silver+
Kikuyu	1,404 sentences	1,055 sentences	Silver
Nepali	40 documents	10 documents	Silver+
Northern Sotho	1,356 sentences	1,125 sentences	Silver
Polish	40 documents	10 documents	Silver+
Somali	605 documents	50 documents	Gold
Swahili	40 documents	10 documents	Silver+
Yoruba	197 documents	50 documents	Gold

- Tools and Reading List
  - http://nlp.cs.rpi.edu/kbp/2017/tools.html
  - http://nlp.cs.rpi.edu/kbp/2017/elreading.html

Window 1 Tri-lingual EDL (part of Cold-Start++ KBP) Participants

Team	Affiliation	Tri-lingual		
		CMN	ENG	SPA
	1st Evaluation Window			
A2KD_Adept	Raytheon BBN Technologies	✓	✓	
ICTCAS_OKN	Institute of Computing Technology, Chinese Academy of Sciences		✓	
ISCAS_Sogou	Institute of Software, Chinese Academy of Sciences & Sogou, Inc.	✓		
SAFT_ISI	USC Information Sciences Institute	✓	✓	✓
STANFORD	Stanford University	✓	✓	✓
TinkerBell	RPI, UIUC, Stanford, Columbia, Cornell, JHU, UPenn	✓	✓	✓
hltcoe	Human Language Technology Center of Excellence	✓	✓	
newbie_mr	Machine Reading Co		✓	

Window 1 Tri-lingual EDL (part of Cold-Start++ KBP)

Performance (Top team = TinkerBell)

Team	NER			NERC			NERLC			KBIDs			CEAFmC+		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
3	**83.2**	**67.3**	**74.4**	**76.8**	**62.2**	**68.8**	**62.6**	**50.7**	**56.0**	**73.1**	**64.9**	**68.8**	**60.7**	**49.1**	**54.3**
13	52.8	54.8	53.8	29.8	30.9	30.3	22.6	23.4	23.0	64.1	46.9	54.2	19.7	20.5	20.1
8	81.7	53.0	64.3	71.7	46.5	56.4	5.5	3.5	4.3	0.0	0.0	0.0	4.8	3.1	3.7
Chinese															
3	84.8	62.9	**72.2**	79.6	59.1	**67.8**	**65.1**	**48.3**	**55.4**	79.9	**64.9**	**71.7**	**64.0**	**47.5**	**54.5**
18	75.0	60.5	67.0	70.0	56.5	62.6	47.8	38.5	42.7	**84.4**	38.7	53.1	46.3	37.4	41.4
13	68.2	47.4	55.9	38.8	26.9	31.8	31.5	21.9	25.8	62.3	44.4	51.8	30.6	21.3	25.1
17	79.8	56.2	66.0	73.9	52.0	61.1	14.7	10.3	12.1	0.0	0.0	0.0	13.9	9.8	11.5
23	56.2	**71.5**	63.0	51.7	**65.9**	57.9	9.9	12.7	11.1	0.0	0.0	0.0	8.9	11.4	10.0
8	**85.4**	50.8	63.7	**81.1**	48.3	60.5	5.0	3.0	3.7	0.0	0.0	0.0	4.6	2.8	3.5
English															
3	77.5	66.7	71.7	71.5	61.5	66.1	**57.9**	49.8	**53.5**	63.6	**68.2**	**65.8**	**54.1**	46.5	**50.1**
18	78.6	79.1	**78.8**	72.6	**73.0**	**72.8**	52.9	**53.2**	53.0	**70.4**	49.8	58.4	48.8	**49.1**	49.0
17	73.0	**79.5**	76.1	66.1	71.9	68.9	23.2	25.3	24.2	0.0	0.0	0.0	21.1	22.9	22.0
19	**90.8**	62.5	74.1	**83.3**	57.3	67.9	26.9	18.5	21.9	0.0	0.0	0.0	23.5	16.2	19.2
13	55.9	70.5	62.4	31.7	39.9	35.3	19.5	24.6	21.8	66.9	50.5	57.6	16.0	20.2	17.9
8	78.5	48.9	60.3	71.3	44.5	54.8	7.8	4.9	6.0	0.0	0.0	0.0	7.0	4.4	5.4
22	51.5	32.9	40.1	29.7	19.0	23.2	5.2	3.3	4.0	0.0	0.0	0.0	4.9	3.1	3.8
Spanish															
3	**86.6**	**74.3**	**80.0**	**78.5**	**67.4**	**72.5**	**64.1**	**55.0**	**59.2**	**76.4**	**62.1**	**68.5**	**62.8**	**53.9**	**58.0**
13	40.9	50.4	45.1	22.7	28.0	25.1	19.9	24.6	22.0	64.0	46.6	53.9	16.2	20.0	17.9
8	84.9	58.7	69.4	63.5	43.9	51.9	5.2	3.6	4.2	0.0	0.0	0.0	4.5	3.1	3.7

Window 2 Tri-lingual EDL Participants (Top team = TAI)

Team	Affiliation	Tri-lingual CMN	ENG	SPA
2089Pacific	Individual		✓	
BUPTTeam	Beijing University of Posts and Telecommunications	✓	✓	✓
Boun	Boğaziči University University		✓	
CMUCS	Language Technologies Institute, Carnegie Mellon University	✓	✓	✓
CRIM	Computer Research Institute of Montreal		✓	
IBM	IBM Research	✓	✓	✓
IRIS	Paul Sabatier University		✓	
NUDT	College of Computer, National University of Defense Technology	✓	✓	✓
RPI_BLENDER	Rensselaer Polytechnic Institute	✓	✓	✓
SUMMA	University College London	✓	✓	✓
TAI	AI platform department of Tencent	✓	✓	✓
UI_CCG	University of Illinois at Urbana Champaign	✓	✓	✓
Ugglan	Lund University	✓	✓	✓
YorkNRM	York University	✓	✓	✓
rise_dcd_zju	College of Computer Science and Technology, Zhejiang University	✓	✓	✓
srcb	Ricoh Software Research Center (Beijing) Co.,Ltd.	✓	✓	

Window 2 Tri-lingual EDL Performance (top team = TAI)

Team	NER			NERC			NERLC			KBIDs			CEAFmC+		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
							Tri-lingual								
1	88.5	71.4	79.0	85.0	68.6	75.9	76.0	**61.3**	**67.8**	78.7	**73.7**	**76.1**	75.4	**60.9**	**67.4**
12	**91.9**	65.0	76.1	**88.2**	62.4	73.1	**81.2**	57.5	67.3	80.5	70.1	75.0	**79.0**	55.9	65.5
7	88.8	64.5	74.7	85.0	61.7	71.5	69.6	50.6	58.6	**81.4**	61.4	70.0	68.7	49.9	57.8
9	83.8	**75.7**	**79.6**	80.8	**72.9**	**76.7**	64.4	58.2	61.1	72.2	64.8	68.3	59.4	53.6	56.4
6	89.4	58.4	70.6	83.0	54.3	65.6	74.9	48.9	59.2	80.1	61.7	69.7	70.9	46.4	56.1
5	87.7	61.5	72.3	81.2	57.0	67.0	71.5	50.1	58.9	68.2	61.2	64.5	67.8	47.5	55.9
2	89.4	60.2	71.9	85.8	57.8	69.1	75.3	50.8	60.7	74.5	65.2	69.5	67.4	45.4	54.3
11	79.5	62.7	70.1	74.3	58.6	65.5	59.8	47.2	52.8	74.1	60.3	66.5	58.3	45.9	51.4
10	88.5	67.7	76.7	85.2	65.3	73.9	68.4	52.4	59.3	76.2	63.1	69.0	58.3	44.6	50.5
14	83.4	51.3	63.5	76.0	46.7	57.9	66.8	41.1	50.9	64.3	47.4	54.5	62.6	38.5	47.6
4	80.2	64.5	71.5	72.5	58.3	64.6	38.9	31.2	34.6	32.0	39.2	35.2	38.3	30.8	34.1

- Is Tri-lingual EDL Solved?
  - Almost perfect linking accuracy for linkable mentions (75.9 vs. 76.1)
  - Almost perfect NIL clustering (67.8 vs. 67.4)
    - perfect name/nominal coreference + cross-doc clustering

Comparison on Three Languages

Best F-score	Extraction	Extraction + Linking	Extraction+Linking +Clustering
English	81.1%	68.4%	66.3%
Chinese	77.3%	71.0%	**70.4%**
Spanish	76.7%	65.0%	64.8%

10 Languages EDL Pilot Participants

- RPI (organizer): 10 languages
- JHU HLT-COE (co-organizer): 5 languages
- IBM: 10 languages

10 Languages EDL Pilot Top Performance

Data	Language	Name Tagging	Name Tagging + Linking
Gold	Chechen	55.4%	52.6%
(from Reflex or	Somali	78.5%	56.0%
LORELEI)	Yoruba	49.5%	35.6%
Silver+	Albanian	75.9%	57.0%
(from Chinese	Kannada	58.4%	44.0%
Rooms)	Nepali	65.0%	50.8%
	Polish	63.4%	45.3%
	Swahili	74.2%	65.3%
Silver (~consistency	Kikuyu	88.7%	88.7%
instead of F)	Northern Sotho	90.8%	85.5%
	All	74.8%	65.9%

- Agreement between Silver+ and Gold is between 72%-85%

What's New and What Works

(Secret Weapons)

Joint Modeling

Turkey's Foreign Minister **Ahmet Davutoglu** greets his supports during an election rally of his ruling AK Party in **Konya**, central Turkey, March 28, 2014.

DBPedia: Justice and Development Party

Properties: Country, headquarter, leaderName, position …

Type Labels: Organization, PoliticalParty, Agent …

DBPedia: Ahmet Davutoğlu

Properties: birthDate, birthPlace, deputy, party, president, successor, religion …

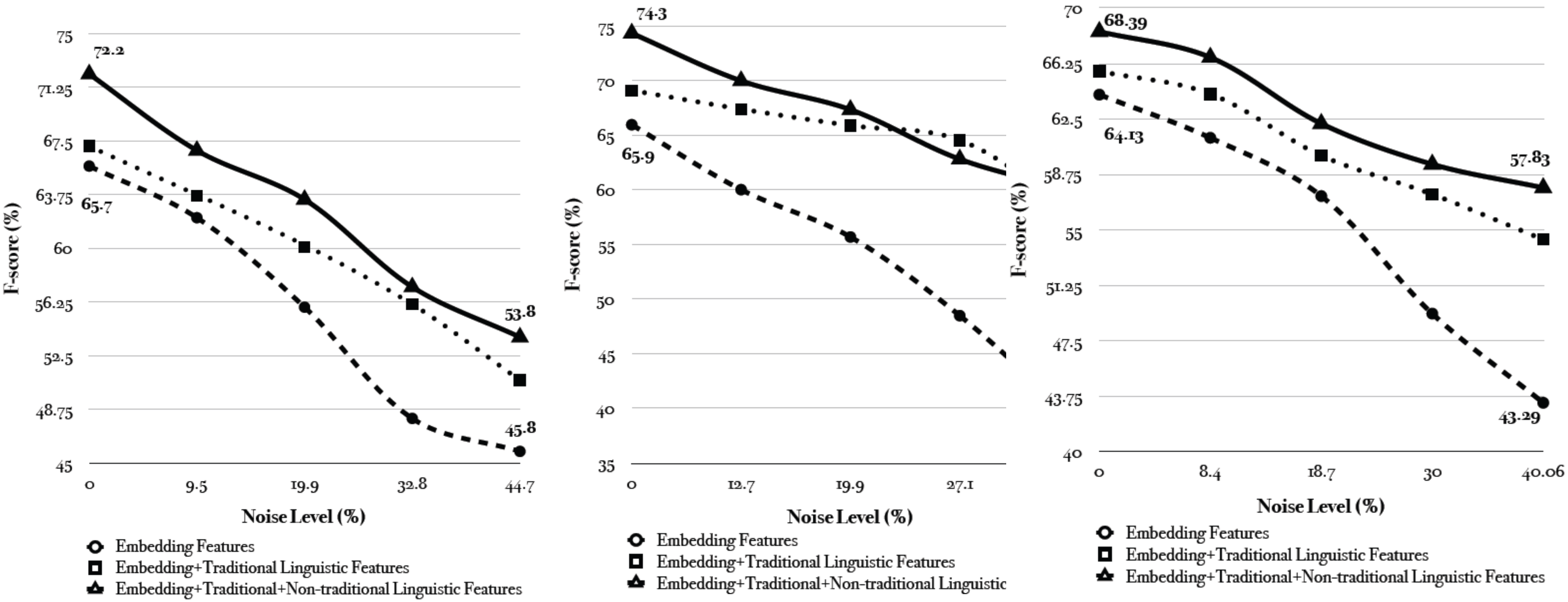
Type Labels: Person, Agent, Politician, Leader, Writer, President, Minister …

- Joint Mention Extraction and Linking (Sil et al., 2013)
  - MSRA team (Luo et al., 2017) designed one single CRFs model for joint name tagging and entity linking and achieved 1.3% name tagging F-score gain
- Joint Word and Entity Embeddings (Cao et al., 2017)
  - CMU (Ma et al., 2017) and RPI (Zhang et al., 2017b)

Return of Supervised Models: Name Tagging

- Rich resources for English, Chinese and Spanish
  - 2009 – 2017 annotations: EDL for 1,500+ documents and EL for 5,000+ query entities
  - ACE, CONLL, OntoNotes, ERE, LORELEI,…
- Supervised models have become popular again
- Name tagging
  - distributional semantic features are more effective than symbol semantic features (Celebi and Ozgur, 2017)
  - combining them significantly enhanced both of the quality and robustness to noise for low-resource languages (Zhang et al., 2017)
- Select the training data which is most similar to the evaluation set (Zhao et al., 2017; Bernier-Colborne et al., 2017)

Incorporate Non-traditional Linguistic Knowledge to make DNN more robust to noise



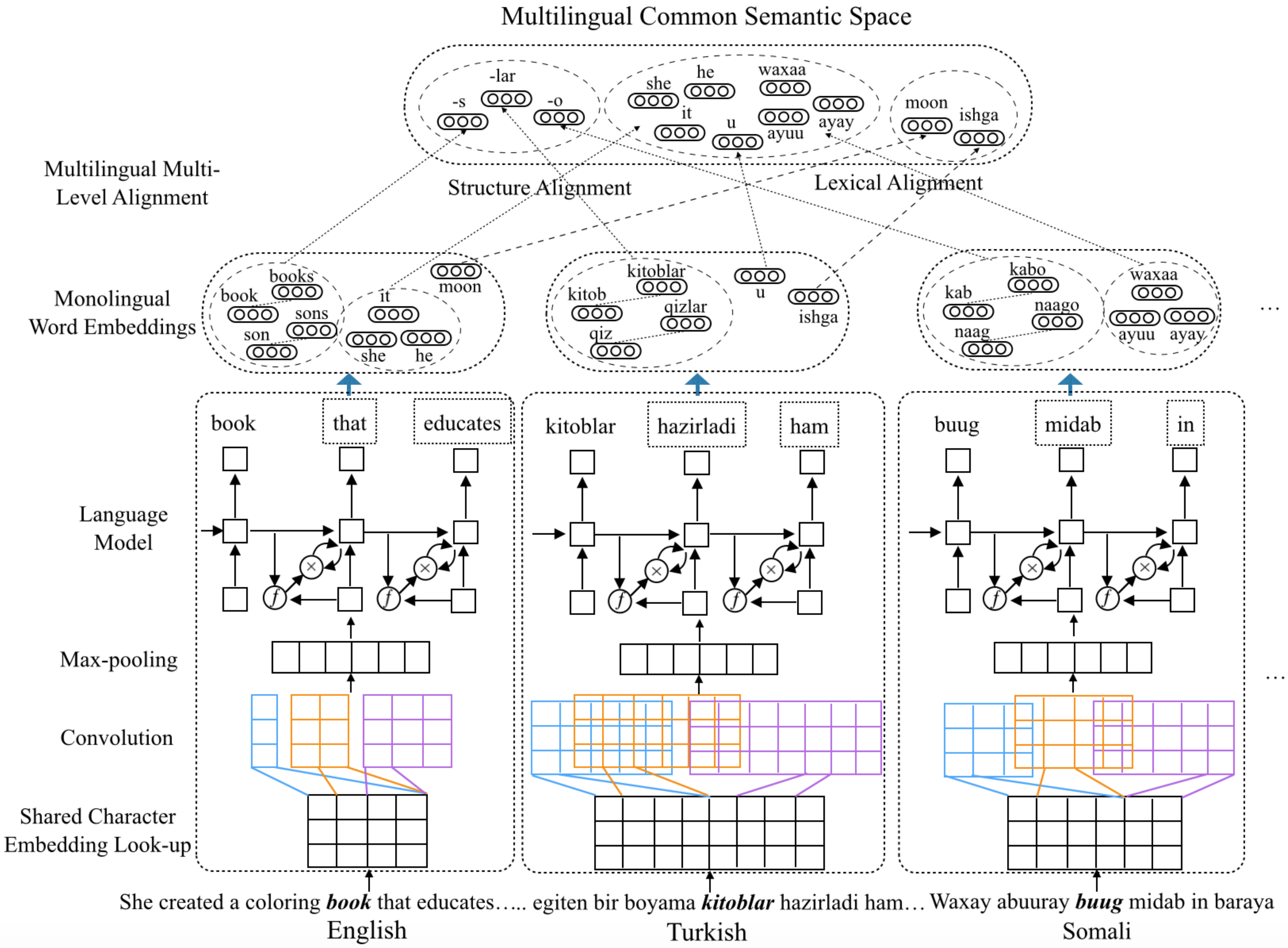
- Zhang et al., 2017

Return of Supervised Models: Entity Linking

- (Sil et al., 2017; Moreno and Grau, 2017; Yang et al., 2017) returned to supervised models to rank candidate entities for entity linking
- The new neural entity linker designed by IBM (Sil et al., 2017) achieved higher entity linking accuracy than state-of-the-art on the KBP2010 data set

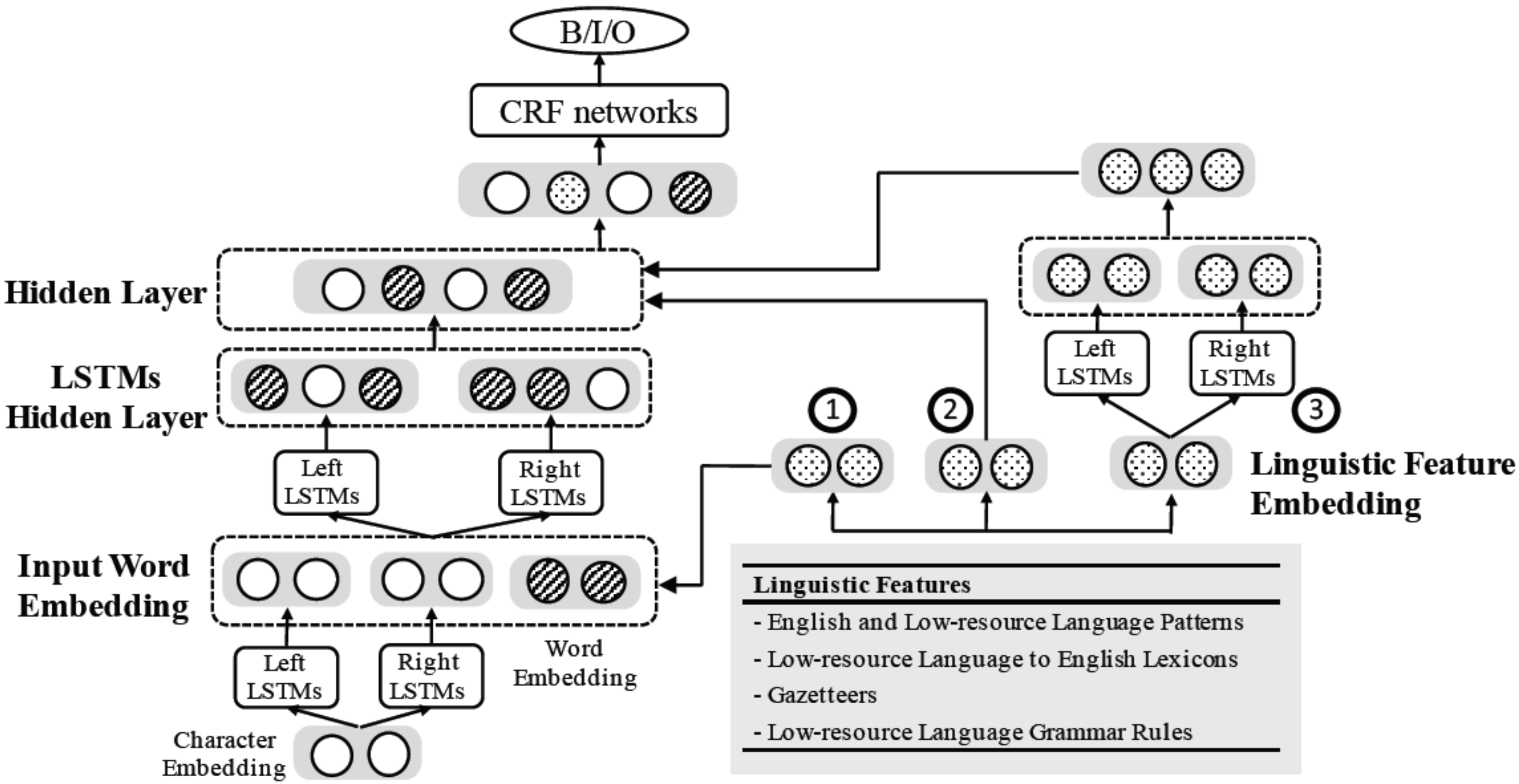
Cross-lingual Common Semantic Space

- Common Space (Zhang et al., 2017)
- Zero-shot Transfer Learning (Sil et al., 2017)



Remaining Challenges

A Typical Neural Name Tagger



Duplicability Problem about DNN

- Many teams (Zhao et al., 2017; Bernier-Colborne et al., 2017; Zhang et al., 2017b; Li et al., 2017; Mendes et al., 2017; Yang et al., 2017) trained this framework
  - the same training data (KBP2015 and KBP2016 EDL corpora)
  - the same set of features (word and entity embeddings)
- Very different results
  - ranked at the 1st, 2nd, 4th, 11th, 15th, 16th, 21st
  - mention extraction F-score gap between the best system and the worst system is about 24%
- Reasons?
  - hyper-parameter tuning?
  - additional training data? dictionaries? embedding learning?
- Solutions
  - Submit and share systems
  - More qualitative analysis

Domain Gap

Name Taggers F-score	Trained from Chinese-Room News	Trained from Wikipedia Markups
Alabanian	75.9%	54.9%
Kannada	58.4%	32.3%
Nepali	65.0%	31.9%
Polish	55.7%	63.4%
Swahili	74.2%	66.4%

- Topic/Domain selection is more important than the size of data
- Tested on news, with ground truth adjudicated from annotations by five annotators through two Chinese Rooms

Glass-Ceiling of Chinese Room

- 72%-85% agreement with Gold-Standard for various languages
- What NIs can do but Non-native speakers cannot:
  - ORGs especially abbreviations, e.g., ኢህወዴግ (Ethiopian People's Liberation Front); ኮብራ (Cobra)
  - Uncommon persons, e.g., ባባ መዳን (Baba Medan)
- Generally low recall

- Reaching the glass ceiling what non-native speakers can understand about foreign languages, difficult to do error analysis and understand remaining challenges
- Need to incorporate language-specific resources and features
- Move human labor from data annotation to interface development to some extent

Background Knowledge Discovery

- Requires deep background knowledge discovery from English Wikipedia and large English corpora: surface lexical / embedding features are not enough
  - *Before 2000, the regional capital of Oromia was* ***Addis Ababa****, also known as ``*__*Finfinne*__*”.*
  - ***Oromo Liberation Fron****t: The armed Oromo units in the Chercher Mountains were adopted as the military wing of the organization, the* ***Oromo Liberation Army*** *or OLA.*
  - ***Jimma Horo*** *may refer to: Jimma Horo,* ***East Welega****, former woreda (district) in East Welega Zone, Oromia Region, Ethiopia; Jimma Horo,* ***Kelem Welega****, current woreda (district) in* ***Kelem Welega Zone****, Oromia Region, Ethiopia*
  - ***Somali*** *(Somali region) !=* ***Somalia*** *!=* ***Somaliland***
    - *The Ethiopian Somali Regional State (Somali: Dawlada Deegaanka Soomaalida Itoobiya) is the easternmost of the nine ethnic divisions (kililoch) of Ethiopia.*
    - *Somalia, officially the Federal Republic of Somalia(Somali: Jamhuuriyadda Federaalka Soomaaliya), is a country located in the Horn of Africa.*
    - *Somaliland (Somali: Somaliland), officially the Republic of Somaliland (Somali: Jamhuuriyadda Somaliland), is a self-declared state internationally recognised as an autonomous region of Somalia.*

Looking Ahead

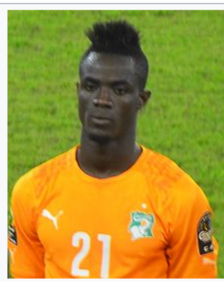
Multi-Media EDL

Jesse Lingard

Mole Valley District

Non-metropolitan district

Eric Bailly



NNUH | Trade Union | Norfolk

contract imposition | save NHS

NHS | Norwich

Junior Doctors Strike

EXIF - 2016:04:26 08:16:17

Geo: Colney, England, United Kingdom

Norfolk and Norwich University Hospital

Norfolk and Norwich University Hospital – NHS Foundation Trust

Multi-Media EDL

- How to build a common cross-media schema?

Speech/Text	Image/Video	Speech/Text	Image/Video
PER.Indefinite	Business_People	FAC.Path	Bridges, Highway, Streets,Tunnel
PER.Individual	Face, Driver,Female_Person	FAC.Airport	Airport,Airport_Or_Airfield,Runway
PER.Group	Backpackers, Officers	FAC.Plant	Power_Plant,Processing_Plant
PER.Individual, FAC.Subarea-Facility	Studio_With_Anchorperson	VEH.Water	Boat_Ship,Canoe,Cigar_Boats, Freighter,Raft,Rowboat,Ship
PER.Individual,WEA	Armed_Person	VEH.Land	Bus,Emergency_Vehicles,Motorcycle
LOC.Water-Body	Beach,Lakes,Oceans,River	VEH.Air	Airplane, Helicopters
LOC.Land-Region-Natural	Mountain,Islands,Valleys	WEA.Projectile	Artillery
FAC.Subarea	Bathroom,Classroom,Court	WEA.Shooting	Machine_Guns, Rifles
FAC.Building-Grounds	Clock_Tower,Shopping_Mall	GPE, ORG	Landmark

- What type of entity mentions should we focus on?

Named pattern: "President Barack"

Riot Police

- How much inference is needed?

NYC?

Streaming Mode

- Perform extraction, linking and clustering at real-time
- Dynamically adjust measures and construct/update KB
- Clustering must be more efficient than agglomerative clustering techniques that require $O(n^2)$ space and time
- Smarter collective inference strategy is required to take advantage of evidence in both local context and global context
- Encourage imitation learning, incremental learning, reinforcement learning

Extended Entity Types

- Extend the number of entity types from five to thousands, so EDL can be utilized to enhance other NLP tasks such as Machine Translation

- 1,000 entity types have clean schema and enough entities in Wikipedia; the English tokens in Wikipedia with these entity types occupy 10% vocabulary

Resources and Evaluation

- Prepare lots of development and test sets in lots of languages, as gold-standard to validate and measure our research progress

- Submit systems instead of results

EDL Systems, Data and Resources

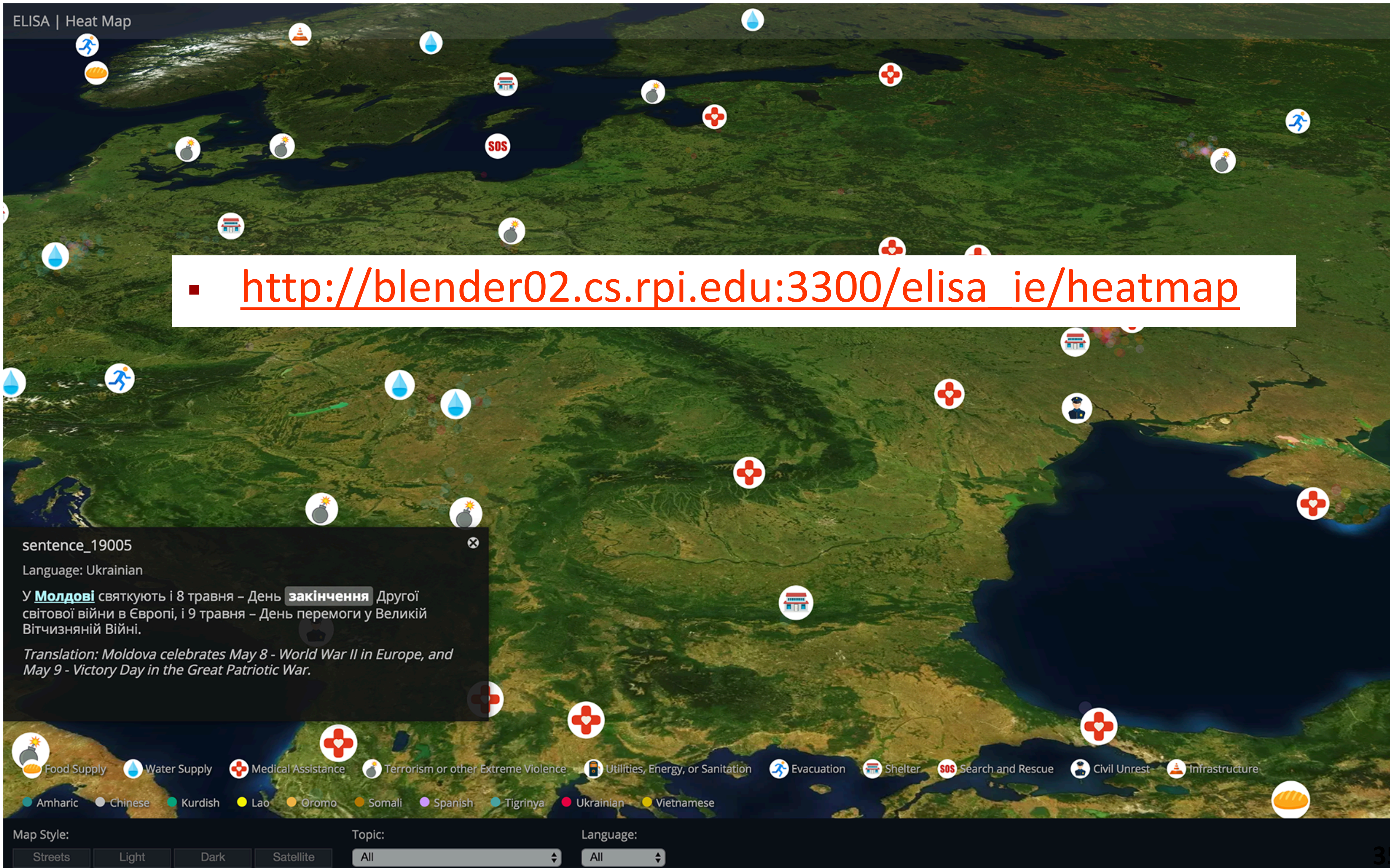
- Resources and Tools
  - http://nlp.cs.rpi.edu/kbp/2017/tools.html

- Re-trainable RPI Cross-lingual EDL Systems for 282 Languages:
  - API: http://blender02.cs.rpi.edu:3300/elisa_ie/api
  - Data, resources and trained models: http://nlp.cs.rpi.edu/wikiann/
  - Demos: http://blender02.cs.rpi.edu:3300/elisa_ie
  - Heatmap demos: http://blender02.cs.rpi.edu:3300/elisa_ie/heatmap

- Share yours!

Thank you for a wonderful decade!

Cross-lingual Entity Discovery and Linking



- http://blender02.cs.rpi.edu:3300/elisa_ie/heatmap

Where We Have Been

Grow with DEFT	2006-2011	2012-2017
Mention Extraction	Human (most)	Automatic
NIL Clustering	None	64 methods
Foreign Languages	Chinese (5%-10% lower than English)	System for 282 languages (Chinese/Spanish comparable to/Outperform English); research toward 3,000 languages
Document Size	-	500 →90,000 documents
Genre	News, web blog	News, Discussion Forum, Web blog, Tweets
Entity Types	PER, GPE, ORG	PER, GPE, ORG, LOC, FAC, hundreds of fine-grained types for typing
Mention Types	Name or all concepts (most)	Name, Nominal, Pronoun (for BeST)
KB	Wikipedia	Freebase → List only
Training Data	20,000 queries (entity mentions)	500 → 0 documents; unsupervised linking comparable to supervised linking
#(Good) Papers	62	110 (new KBP track at ACL); 6 tutorials at top conferences

Technical Term EDL Examples

- P = 69.6%, R = 61.2%, F = 65.1% on English
- Mandarin and Russian Examples

English	Mandarin	Russian
Intermediate value theorem	介值定理	Теорема о промежуточном значении
p-adic number	*p*进数	Р-адичне число
Virtual memory	虚拟内存	Виртуальная память
Nonlinear filter	非线性滤波器	Нелинейный фильтр
Visual odometry	视觉测距	Визуальная одометрия
Wandering set	游荡集	Неблуждающее множество
Photon	光子	Фотон
Support vector machine	支持向量机	Метод опорных векторов
Neuroscience	神经科学	Нейронауки
Heavy water	重水	Тяжёлая вода
Bus (computing)	总线	Шина

Many are Interesting and Useful for MT

Most Challenging Types for MT	# English entities in Wikipedia	Examples
Quantities	7,992	"30 kilometros" to "30 kilometers"
Dates	962,838	"21 enero 2004" to "january 21, 2004"
English Cognates (e.g., technical terms)	20,365	"метод опорных векторов" to "support vector machine"
Specified disaster words		"地震" to "earthquake"
Person Titles	37,722	"Bosh Vazir" to "prime minister"
Colors	27,678	"màu xanh da trời" to "blue"
Holidays	2,358	"день матері" to "mothers day"

Background Knowledge Discovery

- ***EPRDF = OPDO + ANDM + SEPDM + TPLF***
  - *EPRDF: Ethiopian People's Revolutionary Democratic Front, also called **Ehadig**.*
  - *OPDO: Oromo Peoples' Democratic Organization*
  - *ANDM: Amhara National Democratic Movement*
  - *SEPDM: Southern Ethiopian People's Democratic Movement*
  - *TPLF: Tigrayan People's Liberation Front, also called **Weyane** or **Second Weyane**, perhaps because there was a rebellion group called **Woyane/Weyane** in the Tigray province in 1943*
- **Qeerroo** is not an organization although it has its own website:
  - *The overwhelming belief is that its leaders are handpicked by the TPLF puppet-masters, and the new generation of Oromo youth – known as the 'Qeerroo' – have seen that it is business as usual after the latest reform.*
  - *The Qeerroo, also called the Qubee generation, first emerged in 1991 with the participation of the Oromo Liberation Front (OLF) in the transitional government of Ethiopia. In 1992 the Tigrayan-led minority regime pushed the OLF out of government and the activist networks of Qeerroo gradually blossomed as a form of Oromummaa or Oromo nationalism.*
  - *Today the Qeerroo are made up of Oromo youth. These are predominantly students from elementary school to university, organising collective action through social media. It is not clear what kind of relationship exists between the group and the OLF. But the Qeerroo clearly articulate that the OLF should replace the Tigrayan-led regime and recognise the Front as the origin of Oromo nationalism.*

Progress from Window 1 to Window 2

Best F-score	Extraction	Extraction + Linking	Extraction+Linking+Clustering
Window 1	68.8%	56.0%	54.3%
Window 2	76.7%	67.8%	67.4%