

TAC KBP Entity Detection and Linking Collections

Collection	# queries	# docs	# relevant
2014	84	117,132	49
2015	5,344	166	194,512

Corpus

CMR: query is a mention with surrounding context

Chief Warrant Officer Bobby Hall arrived at 11:15 P.M. Friday at MacDill Air Force Base in Tampa, Fla.

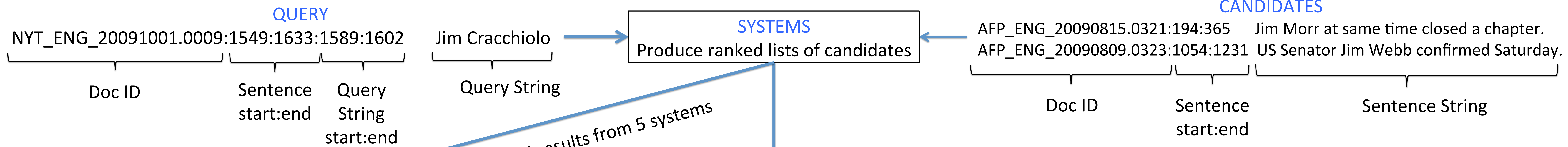
Two weeks after his U.S. Army helicopter dropped from the sky and into an international incident in North Korea, an emotional Bobby Hall flew home late Friday and offered his account of the Dec. 17 crash that led to his 13 days in captivity.

Met at MacDill Air Force Base here by his wife, two children and his parents, Chief Warrant Officer Hall, 28, in a quavering voice, read a statement describing the downing of his helicopter and the death of his co-pilot after the craft veered into the demilitarized zone that separates North and South Korea.

"I really don't know how we ended up across the DMZ," Hall said. "If I knew that, we wouldn't have been there."



<https://github.com/hltcoe/CoreferentMentionRetrieval>



Pooled results from 5 systems

Mechanical Turk Annotation

QUERY MENTION

BANK OF AMERICA SELLS MUTUAL FUND UNIT

"We do n't need two of everything," said Jim Cracchiolo, Ameriprise 's chief executive .

Ameriprise said it hopes to save \$ 130 million to \$ 150 million per year by combining the units .

CANDIDATE MENTION

LAST CALL FOR MOONSHOT

"Jim knew beer and business , and I knew bars , " said Kallman , 50 , who left Boston Beer in 2000 to set out on her own .

" I thought Anheuser-Busch was my competitor , not the federal government . "

How likely is the candidate the same as the query?

very unlikely somewhat unlikely not enough information somewhat likely very likely

- First 3 annotators agreed 61% of the time
- 2 additional annotators for remaining cases
- If <4 of 5 agree marked as "Unknown"

Qrels

Query1	Doc1: Sentence1	Relevant
Query1	Doc1: Sentence2	Unknown
Query1	Doc1: Sentence3	Non-Relevant
Query1	Doc1: Sentence4	Non-Relevant
Query2	Doc1: Sentence5	Unknown
Query2	Doc2: Sentence1	Relevant
Query2	Doc2: Sentence2	Non-Relevant
Query2	Doc2: Sentence3	Relevant
...		

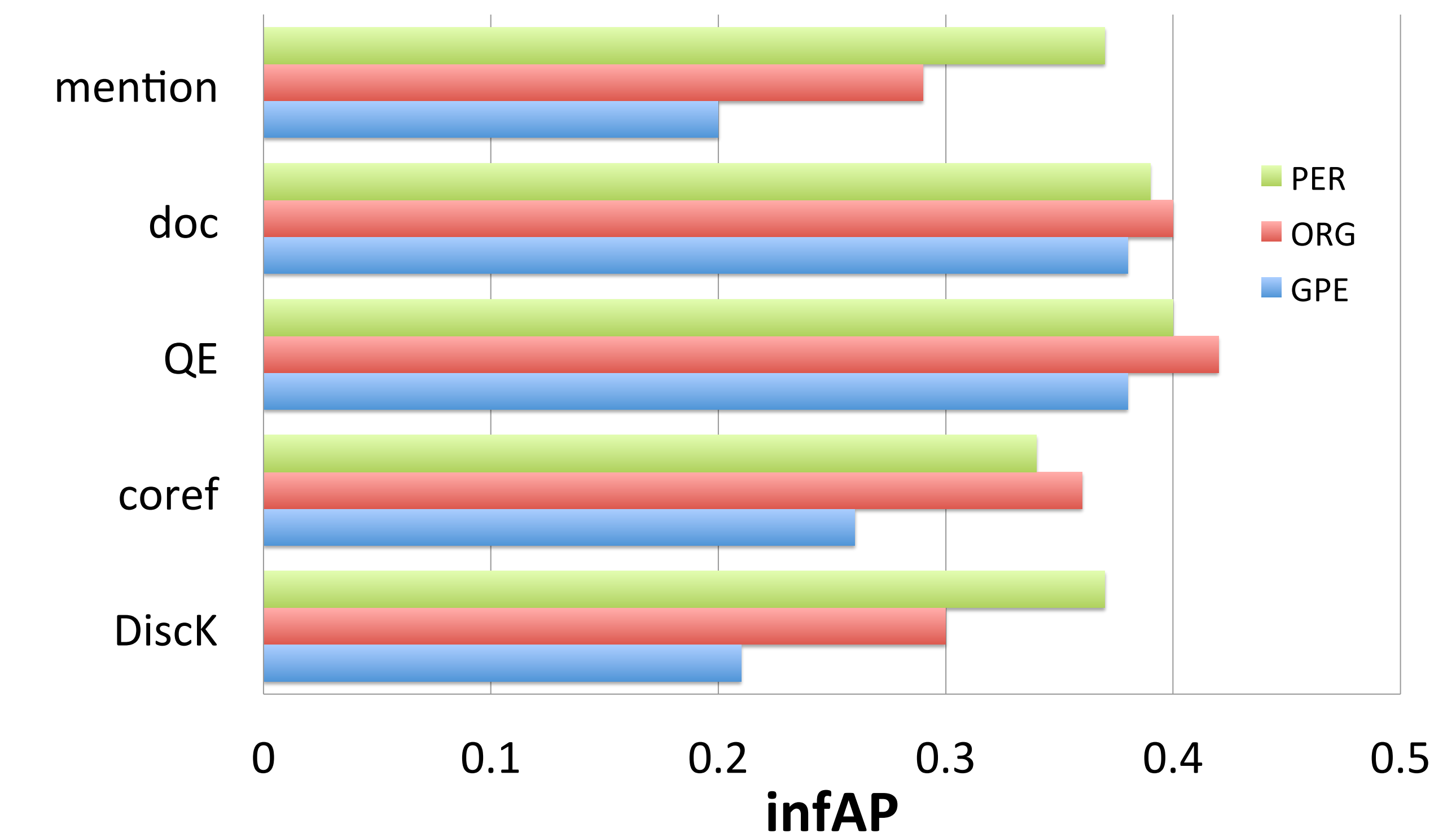
System Results

Query1	Doc6: Sentence5
Query1	Doc9: Sentence4
Query1	Doc2: Sentence2
Query1	Doc6: Sentence1
Query2	Doc4: Sentence1
Query2	Doc9: Sentence3
Query2	Doc5: Sentence4
Query2	Doc6: Sentence4
...	

Baseline Systems

System	Features (weight)
mention	mention words
doc	mention words (9) document words (1)
QE	mention words (0.81) document words (0.09) expansion words (0.1)
coref	document coreference chain
DiscK	mention string, mention words mention type, character trigrams acronyms document words document coreference chain (learned weights)

---all built with Lucene



EVALUATION

Inferred Average Precision (infAP)