Raytheon
BBN Technologies

Cool Stuff at Cold Start: BBN System for TAC 2015

Bonan Min
Marjorie Freedman
Constantine Lignos

Nov 2015

The views expressed are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government. Distribution Statement "A" (Approved for Public Release. Distribution Unlimited.)

This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA) DEFT Program.

Outline

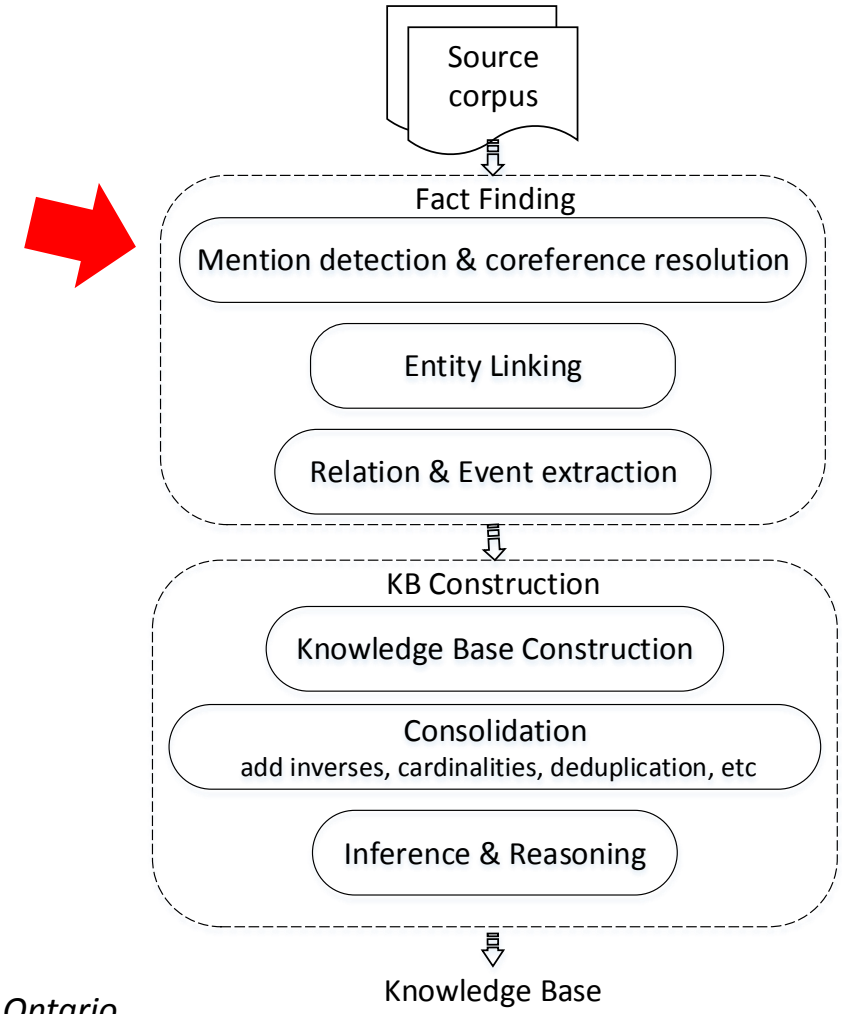
- BBN's 2014 ColdStart System
- What's New
- Entity Discovery & Linking
- Slot Filling
- Experiments and Analysis
- Conclusion

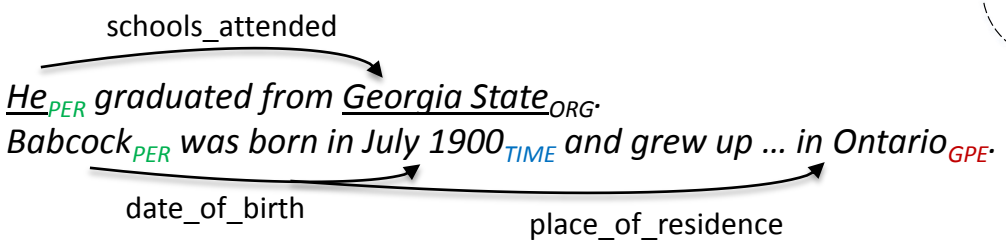
Outline

- **BBN's 2014 ColdStart System**
- What's New
- Entity Discovery & Linking
- Slot Filling
- Experiments and Analysis
- Conclusion

BBN's 2014 ColdStart System

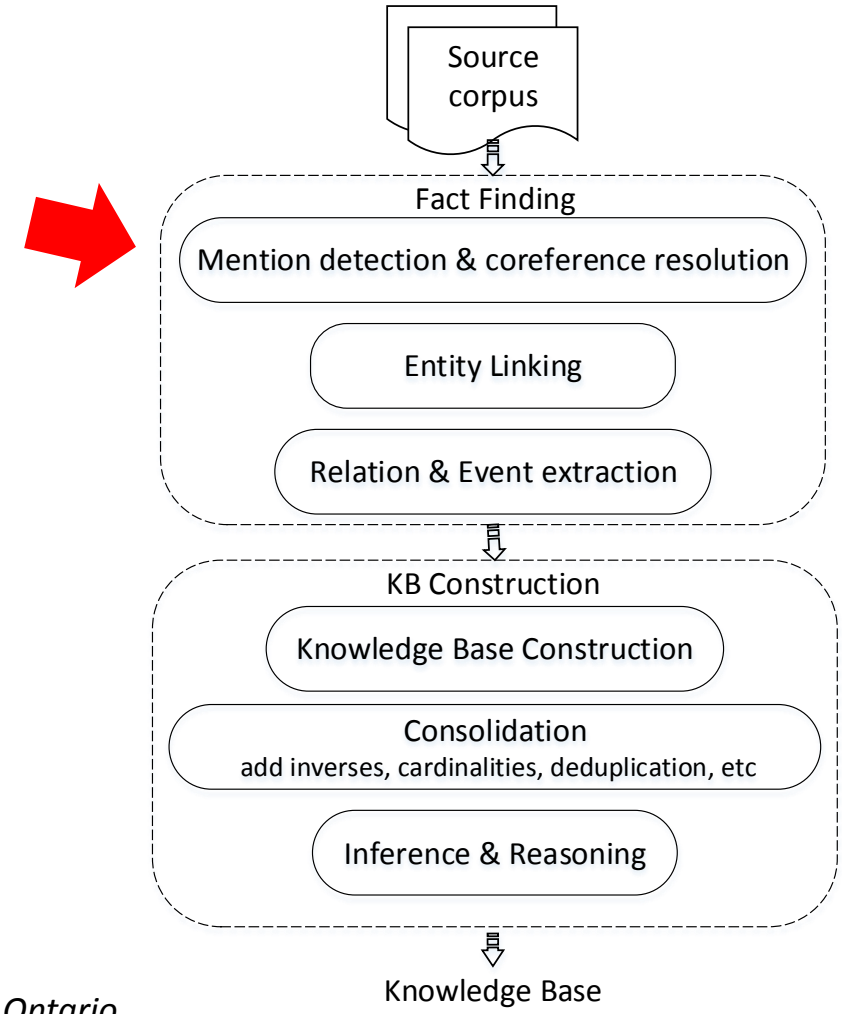
- **Finding local facts**
  - SERIF NLP: syntactic parsing, proposition, etc
  - **Mentions/entities:** SERIF
  - **Relations**
    - SERIF: ACE relations/events
    - Automatically proposed & manually filtered patterns
      - Bootstrap learning
      - Distant supervision
      - Prior-year TAC slot filling annotation
    - Manually created patterns

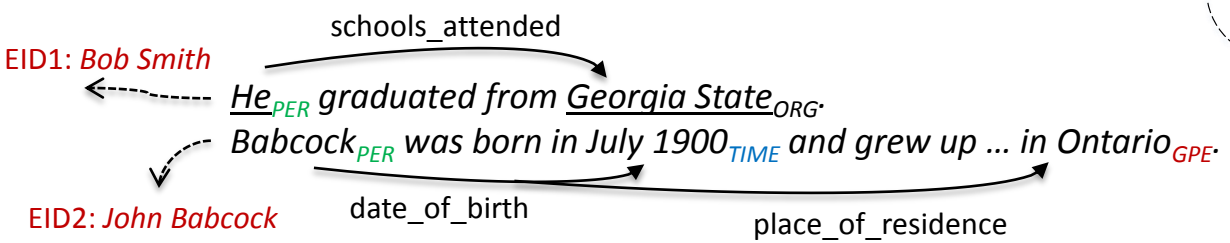




BBN's 2014 ColdStart System

- **Connecting local facts to an entity**
  - Identify mentions with SERIF
  - Link non-names (e.g. he) to named entities with SERIF's within-document coreference
  - Use BBN Actor matching tool to provide corpus-level IDs
    - Matching against a cleaned version of Freebase
    - For name strings that are not matched, clustering based on textual similarity
- **Relations are also deduplicated**





Outline

- BBN's 2014 ColdStart System
- **What's New**
- Entity Discovery & Linking
- Slot Filling
- Experiments and Analysis
- Conclusion

What's New: Outline

- **Entity Discovery and Linking**
  - Address overlinking of names across documents (and within a document)
  - Improve in-doc coreference: coreference ensemble

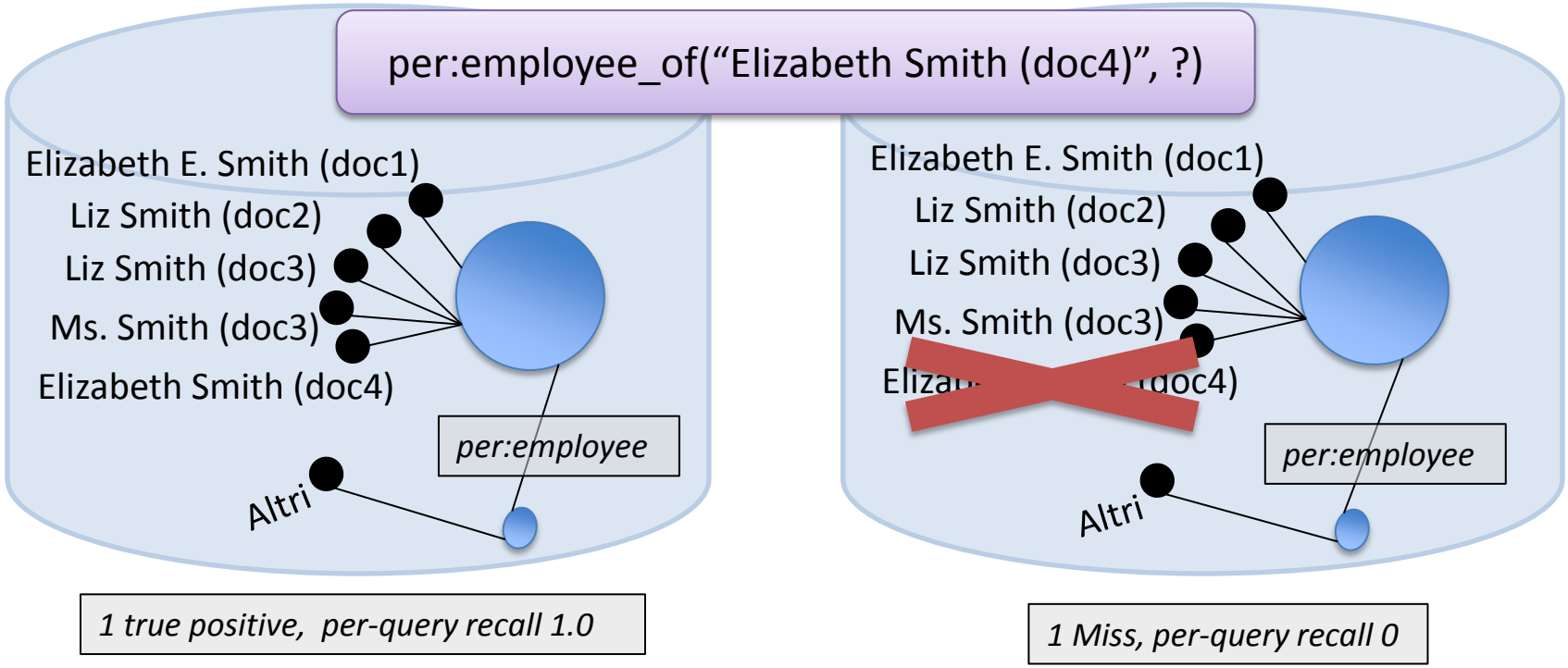
- **Slot Filling**
  - Improve pattern-based extractor
    - Additional patterns proposed from new training sources (CS2014, Rich ERE)
    - Restricted generic patterns
  - Add a statistical relation extraction component

Outline

- BBN's 2014 ColdStart System
- What's New
- **Entity Discovery & Linking**
  - Why focused on Entities?
  - Prevent overlinking
  - Improve with-doc coreference
- Slot Filling
- Experiments and Analysis
- Conclusion

Focus on Entities: Entry Points

- **No credits if missed or miss-typed the entry points**



- **Cold start queries have multiple entry points**
  - *2015 CS adds multiple entry points: In the MAX scoring, a system can answer a query if it finds any of the entry points

Focus on Entities: Cross-Doc Clusters

- **Both within and cross-document linking of names are imperfect**

Within Document Mistakes

Cluster	Text
GPE (Mexico) *Mexico, Mexican …* *New Mexico*	*NYT_ENG_20131105.0194: Vicious **Mexican** drug cartels smuggle cocaine, … through … **New Mexico** to Atlanta … in the United States. … smuggle drugs,… large amounts of cash through Las Cruces, a **New Mexico** city northwest of El Paso*
PER (Ehud Barak) *Ehud Barak, Barak,…* *Barack, Barack Obama*	*44f152b60a8d7bcf65f7b0344764d0d8:**Barak** endorses **Barack**, touts US security support...Defense Minister **Ehud Barak** said Monday night that Barack Obama has been the most supportive president on matters of Israeli security …*

Name Strings Assigned to a Large Cluster

GPE: "U.S." 20,550 mentions
United States (30%), *U.S. (20%), America* (15%), …, *North America*(139 mentions:0.6%), *South America, Central America, … Latin American, Colorado, IDAHO*

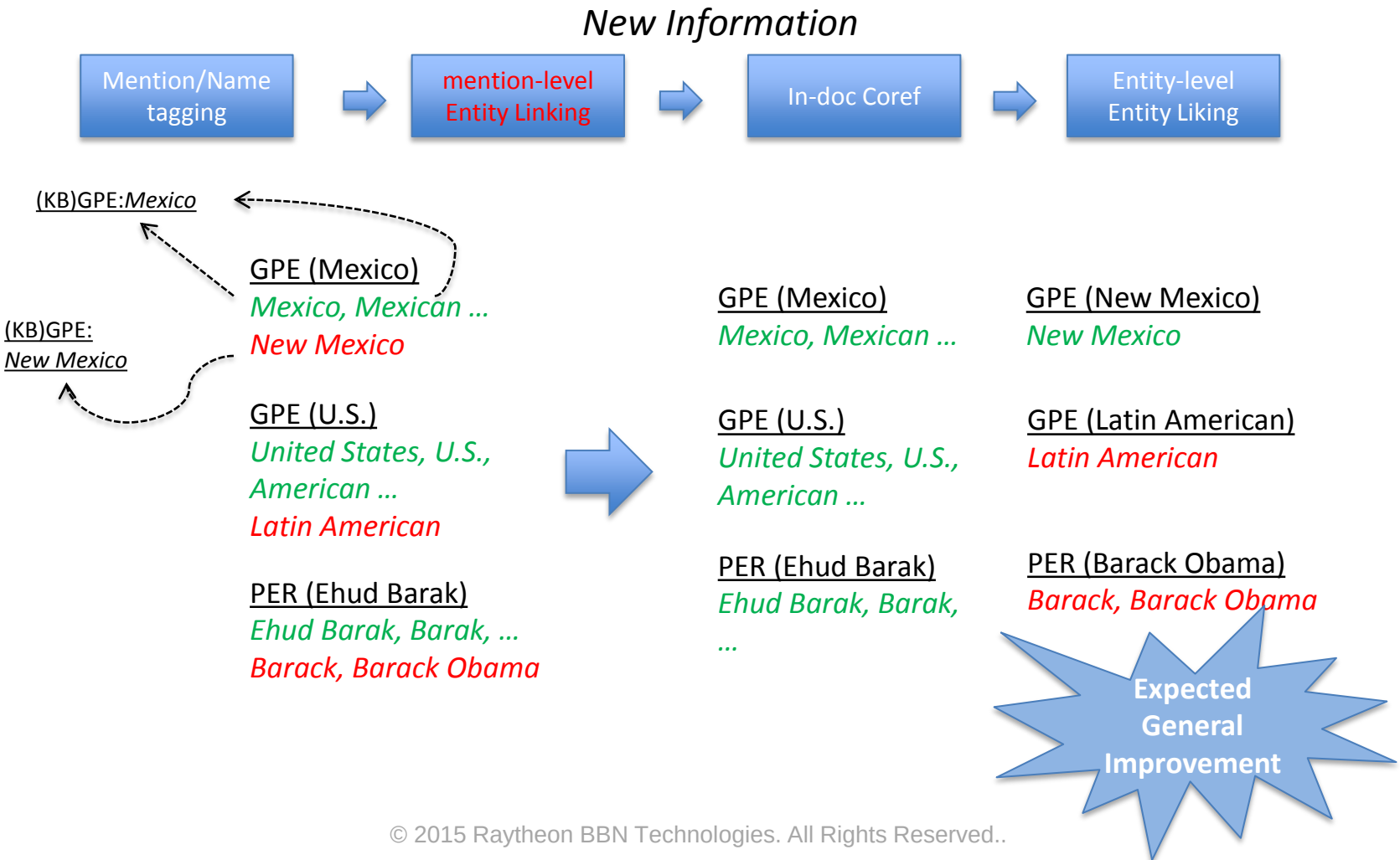
- **BBN's entity discovery mention_ceaf/b_cubed scores exceed other systems, but 'obvious' mistakes persist**
  - Challenging contexts, overlap in name strings, errors in the linking KB…

Outline

- BBN's 2014 ColdStart System
- What's New
- **Entity Discovery & Linking**
  - Why focused on Entities?
  - Prevent overlinking
  - Improve with-doc coreference
- Slot Filling
- Experiments and Analysis
- Conclusion

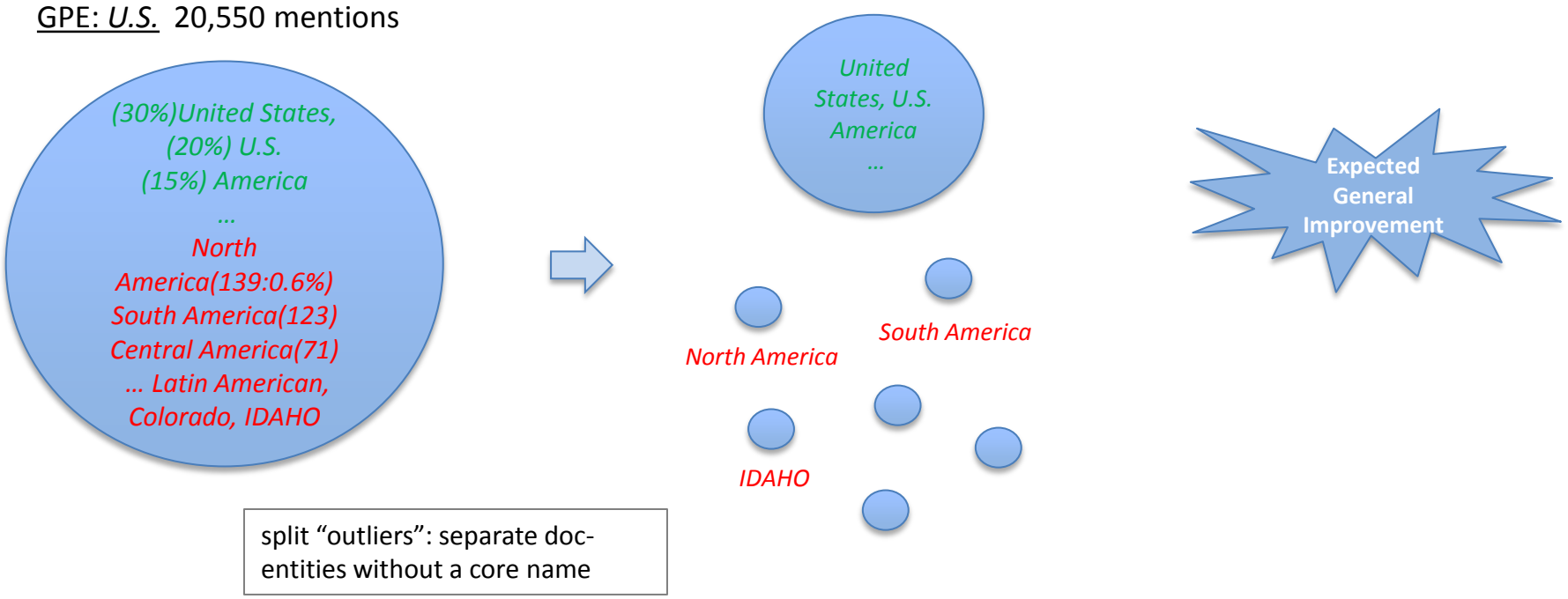
Improving Within Document Coreference

- **Use knowledge (from the Actor DB) about what entities exist in the world to prevent overlinking**



Improving KB Clusters

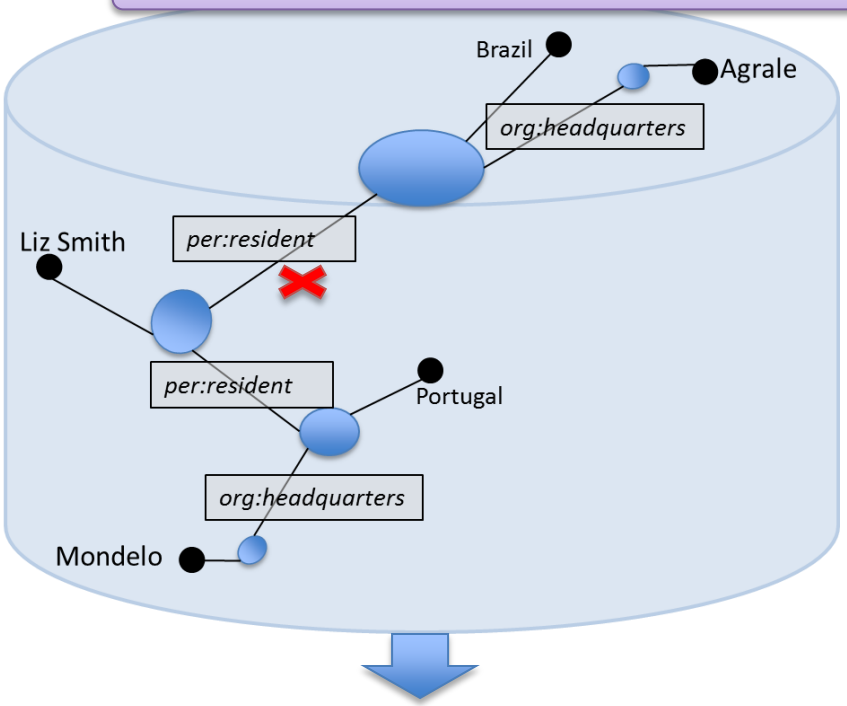
- **EL created clusters on per-document basis**
- **KB consolidation: improve cluster formation at corpus-level**
  - identify core, high frequency names
  - Split outliers into separate entities

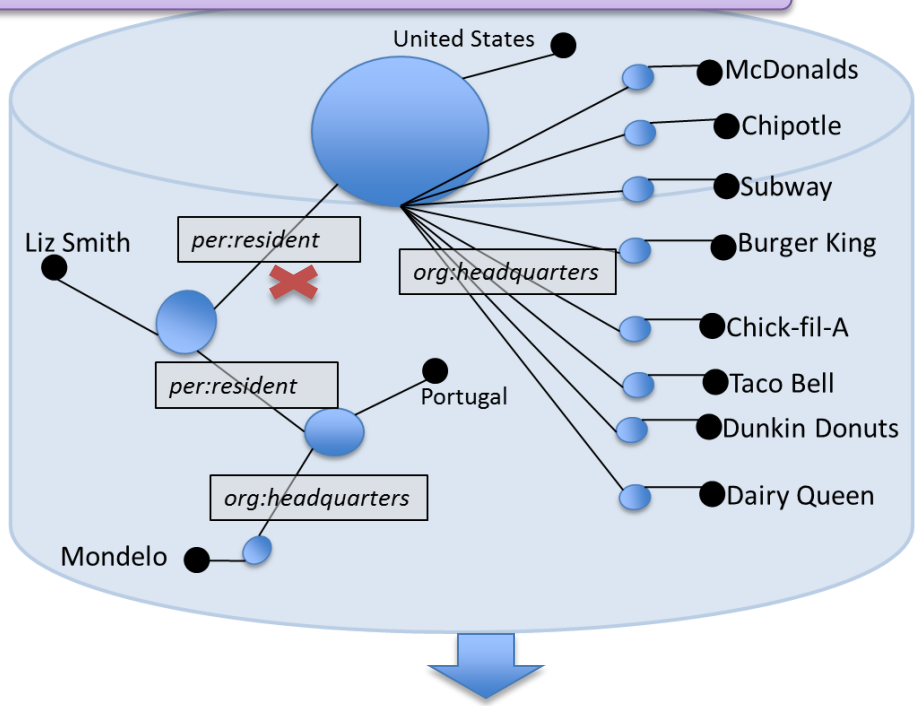


Optimizing for Hop-1

- **Precision for hop-1 queries counts responses as incorrect if the hop-0 was incorrect**
  - This makes entities involved many relations have a high potential risk
  - In 2014, 3 queries led to >50% hop1 FPs

Find organizations that are headquartered in countries where Liz Smith resides.



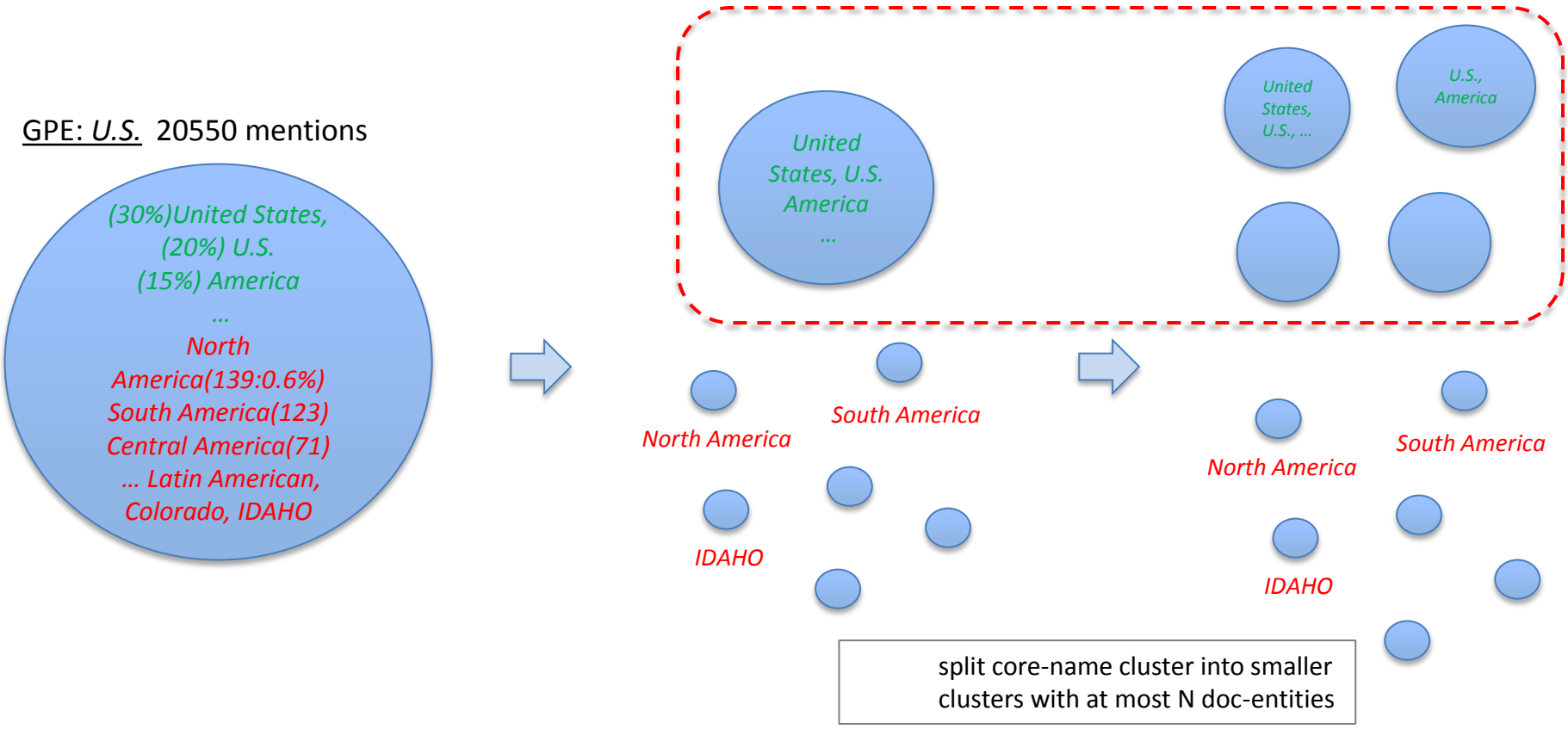


1 incorrect relation in the KB:
Hop 0: 1 True Positive; 1 False Positive; Prec: 0.5
*Hop 1: 1 True Positive**; **1 False Positives; Prec 0.5***

1 incorrect relation in the KB:
Hop 0: 1 True Positive, 1 False Positive; Prec: 0.5
*Hop 1: 1 True Positive, **8 False Positives, Prec: 0.1***

Optimizing for Hop-1 Cont'd

- **Further split very large entities**



Results: Changes on Splitting Entities

- **"BBN": our strong base system**
- **Run-1: + splitting in-doc entities**
- **Run-2: + splitting cross-doc entities (outliers, then very large entities)**
  - Improved precision at hop-0 and hop-1 level, results in BBN's highest "All Hops" score
  - Reduces performance on mention_ceaf and B-Cube
    - optimizing for ED(L) and ColdStart may not be the same

	CS-SF									CS-LDC-MAX								
	Hop0			Hop1			All hops			Hop0			Hop1			All hops		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
BBN	47	31	37	12	13	13	30	24	27	49	37	42	10	18	13	27	30	28
Run-1	47	31	37	12	13	13	30	24	27	49	37	42	10	18	13	27	30	28
Run-2	50	28	36	22	12	16	40	22	28	**52 (+3)**	33(-4)	40(-2)	**20(+10)**	16(-2)	**18(+5)**	**39(+12)**	27(-3)	**32(+4)**

	F1:Mention ceaf	F1: B-Cube
BBN	71	67
Run-2	**69 (-2)**	**66 (-1)**

- **Examples of errors removed**

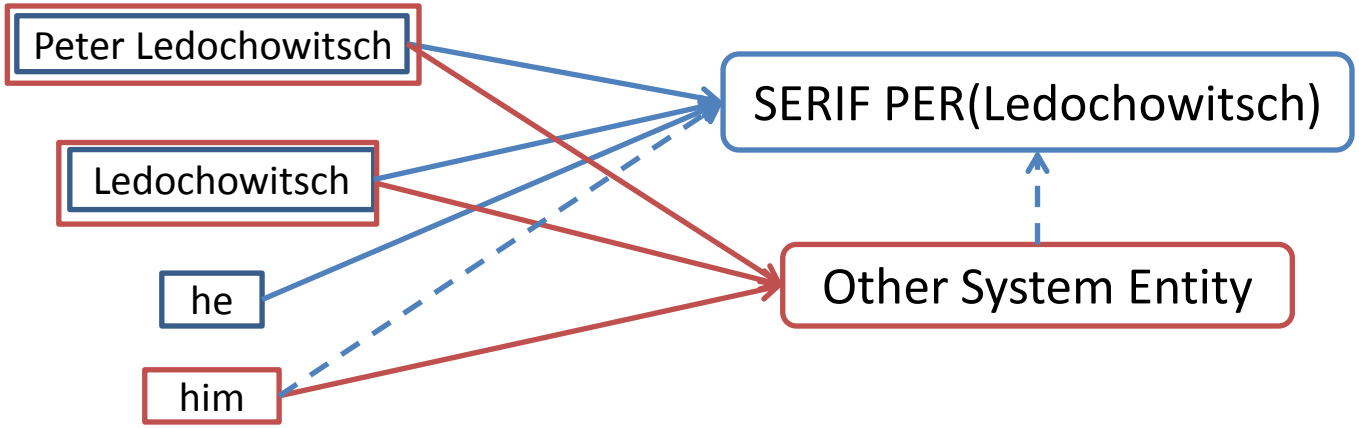
CSSF15_ENG_34c343d331 org:country_of_headquarters "Australia" 79 gpe:residents_of_country
Traditional Anglican Communion

Outline

- BBN's 2014 ColdStart System
- What's New
- **Entity Discovery & Linking**
  - Why focused on Entities?
  - Prevent overlinking
  - Improve with-doc coreference
- Slot Filling
- Experiments and Analysis
- Conclusion

Within-Doc Coreference: Augmentation

- **Improvements to within document coreferences using the Actor DB are limited to name-name coref**
- **Coreference of nominals/pronouns to names also leads to errors**
- **Attempt to merge coref decisions from two systems (+Stanford)**
  - Align secondary system entities to SERIF entities
    - Align using canonical mention if possible
    - Otherwise to the SERIF entity it shares the most mentions with



Results: Coreference Augmentation

		Hop0			Hop1			All hops		
		P	R	F1	P	R	F1	P	R	F1
BBN1	CSSF	46	31	37	12	13	13	29	25	27
	LDC-MAX	49	37	42	10	18	13	27	30	29
BBN3	CSSF	43	31	36	12	13	13	28	25	26
	LDC-MAX	42	37	40	10	18	12	25	31	27

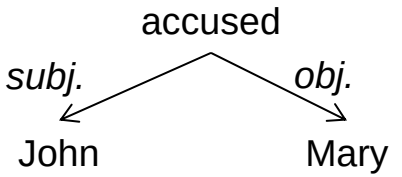
- **24% of added mentions are pronouns**
  - Many are first person pronouns from inside quotes
- **Many useful added mentions are from correctly detecting partial name matches**
  - *Boston Common/the Common*
- **Systems often disagree about desired extents:**
  - *Netherlands/the Netherlands, Boston/Boston's*
  - *[President][Barack Obama]/[President Barack Obama]*
- **Over-linking problems often propagate**
  - Stanford system often links "*X University*" and "*Y University*" together, transferring the error to Serif entity

Outline

- BBN's 2014 ColdStart System
- What's New
- Entity Discovery & Linking
- **Slot Filling**
  - Improve pattern-based extractor
  - Statistical relation extraction
- Experiments and Analysis
- Conclusion

Pattern-based Extractor

- **Proposition patterns: compact, flexible, and accurate representation of many surface-level constructions**



John accused Mary

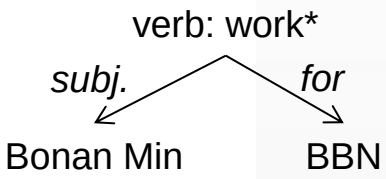
John, a friend of Sheila, accused Mary

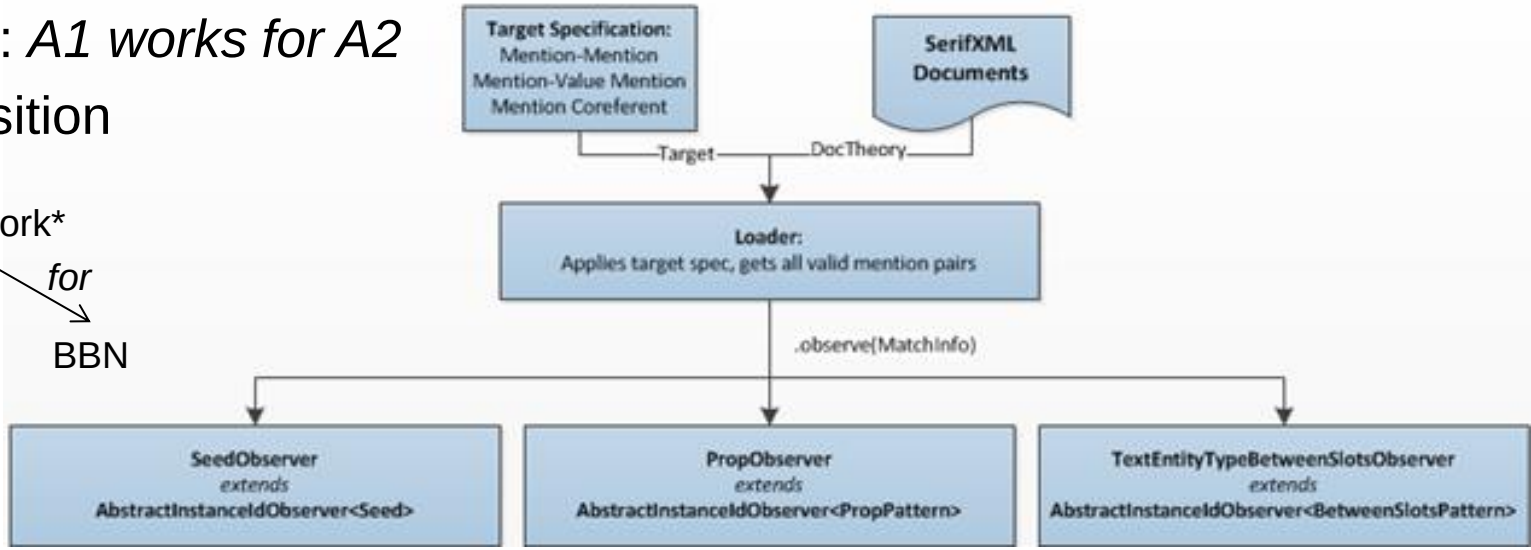
Mary was accused by John

John accused his British friend, Mary

John was sorry to accuse his friend Mary

- **Cold Start patterns**
  - Regex: *A1 works for A2*
  - Proposition





Automatic pattern proposing

Pattern-based Extractor: Improvements

- **Improvement #1: additional patterns proposed from Rich ERE & 2014 CS assessment**

- **Improvement #2: use broad-coverage patterns, e.g., "A1 *of* A2"**
  - per:resident_of: *a longtime* ***resident*** *of* ***San Diego***
    - A1 of A2 ^ A1:PER ^ A2:GPE ^ contain_word(A1, *"resident"*)
  - per:origin: *an* ***immigrant*** *of* ***Mexican*** *origin*
    - A1 of A2 ^ A1:PER ^ A2:Country ^ text_after_A2 ("*origin*")
  - per:employee_of: ***president*** *of* ***Harvard***
    - A1 of A2 ^ A1:PER ^ A2:ORG ^ contain_word(A1, *"president"*)

- **Type of restrictions (automatically collected from context of patterns)**
  - argument type, words, text before/after arguments, etc
- **Apply to broad-coverage patterns**

Results: Improved Pattern-based Extractor

- **BBN 2014: BBN's 2014 Cold Start system**
- **BBN1: higher precision and recall**
  - Added more patterns proposed from the new datasets
  - Added restricted generic patterns

	Hop0			Hop1			All hops		
	P	R	F1	P	R	F1	P	R	F1
BBN2014	45	29	35	10	11	11	28	22	25
BBN1	**47(+2)**	**31(+2)**	**37(+2)**	**12**	**13**	**13**	**30**	**24**	**27**

CS-SF score on 2015 assessment

	Hop0			Hop1			All hops		
	P	R	F1	P	R	F1	P	R	F1
BBN2014	47	34	40	9	15	11	25	28	27
BBN1	**49(+2)**	**37(+3)**	**42(+2)**	**10**	**18**	**13**	**27**	**30**	**28**

CS-LDC-MAX score on 2015 assessment

Outline

- BBN's 2014 ColdStart System
- What's New
- Entity Discovery & Linking
- **Slot Filling**
  - Improve pattern-based extractor
  - Statistical relation extraction
- Experiments and Analysis
- Conclusion

Statistical Relation Extraction

- **Training source**
  - Distant supervision (DS)
    - Freebase pairs → Gigaword
  - TAC: query → mention in reference doc
  - Rich ERE: mention-level annotation

Dataset	# Examples
SF2013	5268
SF2012	3825
SF2014	3549
CS2013	1689
CS2014	1350

- **Features**

Strict Model

Feature	Example
Regex+ArgType	*PER:smith_work*_for_ORG:company*
Prop+ArgType	*verb:works[PER:smith:<sub>][ORG:company:for]*
Head words	*smith; company; smith-company*
Bag of words	*A1: John; smith A2: company between: works before/after: Amazon*
Entity Type	*A1:PER; A2:ORG; A1-A2-PER-ORG*

Features for "*John Smith works for the eCommerce company Amazon.*"

Statistical Relation Extraction

- **Train MaxEnt models on relation mentions**

ERE		TAC		DS	
per_employee_or_member_of	896	per_employee_or_member_of	1101	per_origin	1938221
per_religion_or_origin	368	per_place_of_residence	595	per_employee_or_member_of	753372
per_place_of_residence	317	per_title	579	per_place_of_birth	272280
per_title	279	org_place_of_headquarters	257	per_place_of_residence	268838
org_parents	209	per_spouse	155	org_place_of_headquarters	153364
		per_origin	143	org_parents	67980
		per_place_of_death	134	per_spouse	46063
		org_parents	86	per_schools_attended	39046
		per_schools_attended	79	org_founded_by	34541
				per_siblings	5949

# of positive examples in the training datasets.
*went through feature-vector deduplication

Results: Statistical Relation Extraction

	Hop0			Hop1			All hops		
	P	R	F1	P	R	F1	P	R	F1
BBN1	47	31	37	12	13	13	30	24	27
BBN4	**40(-7)**	**35(+4)**	**37**	**12**	**18(+5)**	**14(+1)**	**26(-4)**	**29(+5)**	**27**

Added Trained models (BBN4) into the base system (BBN1)

Slot	BBN1	BBN4	Statistical RE	
			# added	% added
ALL	236264	326130	73835	31.25%
per:employee_or_member_of	39268	46488	7220	18.39%
org:subsidiaries	7931	11902	3971	50.07%
per:countries_of_residence	6428	17485	5109	79.48%
per:statesorprovinces_of_residence	2901	8136	1386	47.78%
per:cities_of_residence	7477	10503	3026	40.47%
org:country_of_headquarters	2820	6662	671	23.79%
org:stateorprovince_of_headquarters	1162	3936	292	25.13%
org:city_of_headquarters	4428	4993	565	12.76%
per:schools_attended	1336	3161	1825	136.60%
per:stateorprovince_of_death	59	147	7	11.86%

Relations added by the statistical relation models

Results: Statistical Relation Extraction

- Traned model with each datasets (evaluated on CS2015 & CS2014)

	Hop0		
	P	R	F1
DS	28	9	14
ERE	43	4	8
TAC	51	2	5

Trained model with each dataset only.
High quality features, CS 2015, STRING_CASE, CS-SF score

	Hop0		
	P	R	F1
DS	46	16	23
ERE	52	6	11
TAC	62	7	12

Trained model with each dataset only.
High quality features, CS 2014

- More fine-grained features helped; but with cost of precision)

Feature	Example
Regex+ArgType	*PER:smith_work*_for_ORG:company*
Prop+ArgType	*verb:works[PER:smith:<sub>][ORG:company:for]*
Head words	*smith; company; smith-company*
Bag of words	*A1: John; smith A2: company between: works before/after: Amazon*
Entity Type	*A1:PER; A2:ORG; A1-A2-PER-ORG*

Features for "*John Smith works for the eCommerce company Amazon.*"

	Hop0		
	P	R	F1
TAC	61 *	19	29
TAC+ERE	56 *	27	36

Trained model with datasets;
also Added low-quality features, CS 2014
* Precision numbers are not quite useful since this is post-assessment evaluation

Outline

- BBN's 2014 ColdStart System
- What's New
- Entity Discovery & Linking
- Slot Filling
- **Experiments and Analysis**
- Conclusion

Experiments – official runs

- BBN1: base system
- BBN2: + split of cross-doc entities
- BBN3: + split doc entities + coreference ensemble
- BBN4: + statistical RE + inference
- BBN5: + nested names

	BBN1	BBN2	BBN 3	BBN 4	BBN5
Improved patterns	√	√	√	√	√
Split cross-doc entities		√			
Split in-doc entities			√		
Coreference ensemble			√		
Statistical RE				√	
Inference				√	
Nested names					√

	CS-SF									CS-LDC-MAX								
	Hop0			Hop1			All hops			Hop0			Hop1			All hops		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
BBN1	46	31	**37**	12	13	13	29	25	27	49	37	**42**	10	18	13	27	30	29
BBN2	**49**	28	36	**23**	13	**16**	**40**	22	**29**	**50**	33	40	**21**	17	**19**	**39**	27	**32**
BBN3	43	31	36	12	13	13	28	25	26	42	37	40	10	18	12	25	31	27
BBN4	37	**36**	36	12	**19**	15	25	**30**	27	40	**42**	41	10	**25**	14	23	**36**	28
BBN5	45	29	36	12	14	13	28	23	26	49	35	41	11	21	14	26	30	28

Official scores on the preliminary assessments

Experiments: What Helped

- BBN1: base system (BBN 2014 + more training data + restricted generic patterns)
- BBN2: BBN1 + split of cross-doc entities
- BBN3: BBN1 + split doc entities + coreference ensemble
- BBN4: BBN1 + statistical RE + inference
- BBN5: BBN1 + with finding nested names e.g., identify "*US*" in mention "*US Army*".

Improved patterns

split cross-doc entity

Statistical RE & inference

	Hop0			Hop1			All hops		
	P	R	F1	P	R	F1	P	R	F1
BBN2014	45	29	35	10	11	11	28	22	25
BBN1	47	31	37	12	13	13	30	24	**27 (+2)**
BBN2	50	28	36	**22**	12	16	40	22	**28 (+3)**
BBN3	47	31	37	12	13	13	30	24	27
BBN4	40	**35**	37	12	18	14	26	29	27
BBN5	47	29	36	10	11	10	29	22	25

Scores on the preliminary assessments. * BBN Running Scoring (not NIST reported Scores); use post-hoc flag for all

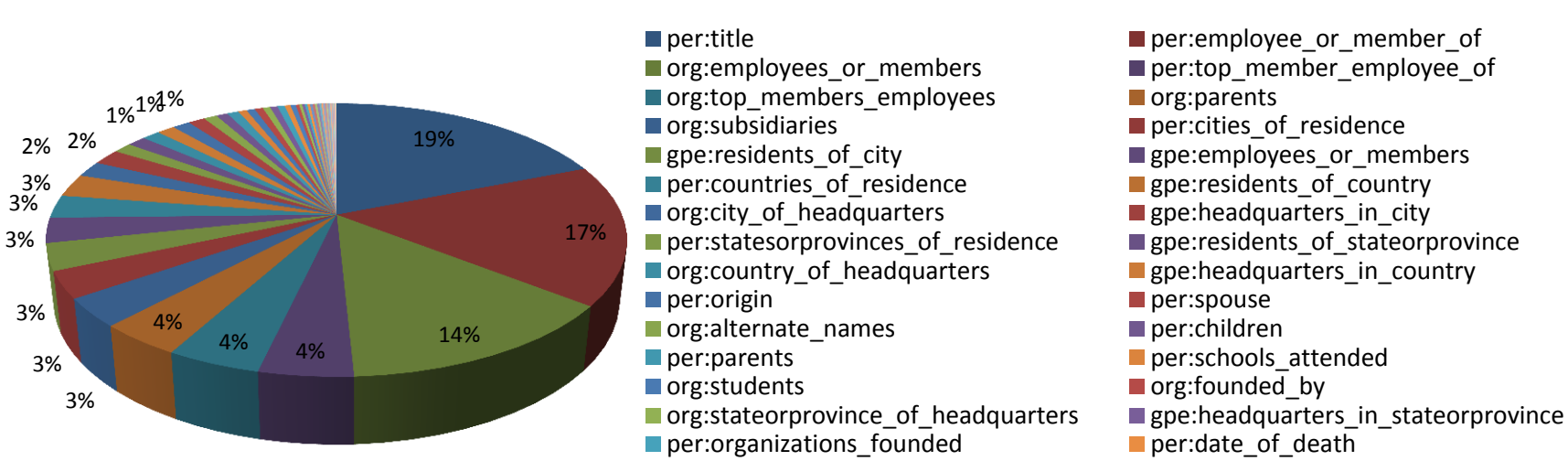
Overall Stats

split cross-doc entity

Finding nested names

	#entities	PER	ORG	GPE
BBN1	210801	124700	67966	18135
BBN2	217885 (+3.3%)	125031	69525	23329 (+28.6%)
BBN3	210812	124702	67974	18136
BBN4	210801	124700	67966	18135
BBN5	214298(+1.6%)	127771(+2.4%)	67213	19314(6.5%)

entities in submissions



Distribution of slots in BBN1

Conclusion: Lessons Learnt

- **Prevent overlinking of entities helped**
- **More and better pattern improves recall (and precision)**
  - More patterns from additional training data
  - Restricted generic patterns
- **Statistical relation extraction boost recall at the cost of precision**

Thanks!