

KBP 2017 Cold Start

KB Construction and Slot Filling

Hoa Dang Shahzad Rajput

U.S. National Institute of Standards and Technology

TAC 2017

- 1 Introduction
 - Task Variants
 - Changes in 2017
- 2 SF/KB Evaluation
 - Definitions
 - Queries
 - Participants
 - Results
- 3 Conclusion

- 1 Introduction
 - Task Variants
 - Changes in 2017
- 2 SF/KB Evaluation
 - Definitions
 - Queries
 - Participants
 - Results
- 3 Conclusion

- Knowledge Base Construction ($K=Many$) - Queries were not provided
 - ED Evaluation
 - Event Argument Evaluation
 - Event Nugget Evaluation
 - BeST Evaluation
 - SF Evaluation ($K=1,3$)
- Slot Filling ($K=1$) - Queries involving SF slots were provided

Task Variants

Queries - Example # 1

LDC Query

```
<query id='CS17_9999'>
  <mentions>
    <mention>
      <name>June McCarthy</name>
      <docid>ENG_142</docid>
      <beg>16931</beg>
      <end>16943</end>
    </mention>
    <mention>
      <name>Junio McCarthy</name>
      <docid>SPA_142</docid>
      <beg>2863</beg>
      <end>2869</end>
    </mention>
  </mentions>
  <enttype>per</enttype>
  <nodeid>per_049</nodeid>
  <slot0>per:children</slot0>
  <slot1>per:age</slot1>
</query>
```

SF Query

```
<query id='CSSF17_ENG_abcabdefde'>
  <name>June McCarthy</name>
  <docid>ENG_142</docid>
  <beg>16931</beg>
  <end>16943</end>
  <enttype>PER</enttype>
  <slot>per:children</slot>
  <slot0>per:children</slot0>
  <slot1>per:age</slot1>
</query>
<query id='CSSF17_SPA_defdeabcab'>
  <name>Junio McCarthy</name>
  <docid>SPA_142</docid>
  <beg>2863</beg>
  <end>2869</end>
  <enttype>PER</enttype>
  <slot>per:children</slot>
  <slot0>per:children</slot0>
  <slot1>per:age</slot1>
</query>
```

Task Variants

Queries - Example # 2

LDC Query

```
<query id='CS17_9999'>
  <mentions>
    <mention>
      <name>June McCarthy</name>
      <docid>ENG_142</docid>
      <beg>16931</beg>
      <end>16943</end>
    </mention>
    <mention>
      <name>Junio McCarthy</name>
      <docid>SPA_142</docid>
      <beg>2863</beg>
      <end>2869</end>
    </mention>
  </mentions>
  <enttype>per</enttype>
  <nodeid>per_049</nodeid>
  <slot0>per:contact.meet_entity</slot0>
</query>
```

SF Query

```
<query id='CSSF17_ENG_abcabdefde'>
  <name>June McCarthy</name>
  <docid>ENG_142</docid>
  <beg>16931</beg>
  <end>16943</end>
  <enttype>PER</enttype>
  <slot>per:contact.meet_entity</slot>
  <slot0>per:contact.meet_entity</slot0>
</query>

<query id='CSSF17_SPA_defdeabcab'>
  <name>Junio McCarthy</name>
  <docid>SPA_142</docid>
  <beg>2863</beg>
  <end>2869</end>
  <enttype>PER</enttype>
  <slot>per:contact.meet_entity</slot>
  <slot0>per:contact.meet_entity</slot0>
</query>
```

Task Variants

Queries - Example # 3

LDC Query

```
<query id='CS17_9999'>
  <mentions>
    <mention>
      <name>June McCarthy</name>
      <docid>ENG_142</docid>
      <beg>16931</beg>
      <end>16943</end>
    </mention>
    <mention>
      <name>Junio McCarthy</name>
      <docid>SPA_142</docid>
      <beg>2863</beg>
      <end>2869</end>
    </mention>
  </mentions>
  <enttype>per</enttype>
  <nodeid>per_049</nodeid>
  <slot0>per:children</slot0>
  <slot1>per:contact.meet_entity</slot1>
</query>
```

SF Query

```
<query id='CSSF17_ENG_abcabdefde'>
  <name>June McCarthy</name>
  <docid>ENG_142</docid>
  <beg>16931</beg>
  <end>16943</end>
  <enttype>PER</enttype>
  <slot>per:children</slot>
  <slot0>per:children</slot0>
  <slot1>per:contact.meet_entity</slot1>
</query>
<query id='CSSF17_SPA_defdeabcab'>
  <name>Junio McCarthy</name>
  <docid>SPA_142</docid>
  <beg>2863</beg>
  <end>2869</end>
  <enttype>PER</enttype>
  <slot>per:children</slot>
  <slot0>per:children</slot0>
  <slot1>per:contact.meet_entity</slot1>
</query>
```

Task Variants

Queries - Example # 4

LDC Query

```
<query id='CS17_9999'>
  <mentions>
    <mention>
      <name>June McCarthy</name>
      <docid>ENG_142</docid>
      <beg>16931</beg>
      <end>16943</end>
    </mention>
    <mention>
      <name>Junio McCarthy</name>
      <docid>SPA_142</docid>
      <beg>2863</beg>
      <end>2869</end>
    </mention>
  </mentions>
  <enttype>per</enttype>
  <nodeid>per_049</nodeid>
  <slot0>per:is_liked_by</slot0>
  <slot1>per:age</slot1>
</query>
```

SF Query

```
<query id='CSSF17_ENG_abcabdefde'>
  <name>June McCarthy</name>
  <docid>ENG_142</docid>
  <beg>16931</beg>
  <end>16943</end>
  <enttype>PER</enttype>
  <slot>per:is_liked_by</slot>
  <slot0>per:is_liked_by</slot0>
  <slot1>per:age</slot1>
</query>
<query id='CSSF17_SPA_defdeabcab'>
  <name>Junio McCarthy</name>
  <docid>SPA_142</docid>
  <beg>2863</beg>
  <end>2869</end>
  <enttype>PER</enttype>
  <slot>per:is_liked_by</slot>
  <slot0>per:is_liked_by</slot0>
  <slot1>per:age</slot1>
</query>
```


Task Variants

Example#01: KB → EDL

Knowledge Base

```
:e4 type PER
:e4 mention "Bart" Doc726:37-40
:e4 nominal_mention "brother" Doc726:15-21
```

EDL component

```
EKB :e4_M1 Bart Doc726:37-40 NIL_0001 PER NAM 1.0
EKB :e4_M2 brother Doc726:15-21 NIL_0001 PER NOM 1.0
```

Task Variants

Example#02: KB → EAL

Knowledge Base

```
:en1 type GPE
:en1 mention "Springfield" Doc726:402-412 1.0

:ev1 type CONFLICT.ATTACK
:ev1 mention.actual "bombing" Doc726:418-424 1.0
:ev1 conflict.attack:place.actual :en1 Doc726:261-431;Doc726:402-407;NIL 1.0
...
```

EAL component (arguments)

```
EKB:ev1.1 Doc726 Conflict.Attack Place Springfield 402-407 261-431 402-407 NIL Actual 1.0
...
```

EAL component (linking)

```
EKB:ev1.Doc726 EKB:ev1.1 ...
```

EAL component (corpuslinking)

```
EKB:ev1 EKB:ev1.Doc726 ...
```

Task Variants

Example#03: KB → EN

Knowledge Base

```
:ev1 type CONFLICT.ATTACK  
:ev1 mention.actual "bombing" Doc726:418-424 1.0
```

EN component

```
#BeginOfDocument Doc726  
Doc726 E1 418,424 bombing Conflict_Attack Actual  
#EndOfDocument
```

Task Variants

Example#04: KB → BeST

Knowledge Base

```
:en1 type GPE
:en1 mention "Springfield" Doc726:402-412 1.0
:en4 type PER
:en4 mention "Bart" Doc726:37-40
:en4 per:dislikes :en1 Doc726:422-427 1.0
```

BeST component

```
<?xml version="1.0" encoding="UTF-8"?>
<belief_sentiment_doc id="Doc726">
  <sentiment_annotations>
    <entities>
      <entity ere_id=":en1.500001" offset="402" length="11">
        <text>Springfield</text>
        <sentiments>
          <sentiment polarity="neg" sarcasm="no" confidence="1.0">
            <source ere_id=":en4.500001" offset="37" length="4">Bart</source>
          </sentiment>
        </sentiments>
      </entity>
    </entities>
  </sentiment_annotations>
</belief_sentiment_doc>
```

Task Variants

Example#05: KB → Slot-filling

Knowledge Base

```
:en7 type PER
:en7 canonical_mention "Bart" Doc726:434-437 1.0

:en10 type PER
:en10 mention "Lisa" Doc726:684-687 1.0

:en10 per:siblings :en7 Doc726:561-577 0.7
```

SF Query

```
<query id="CS17_ENG_54d83ec119">
  <name>Lisa</name>
  <docid>Doc726</docid>
  <beg>684</beg>
  <end>687</end>
  <enttype>per</enttype>
  <slot0>per:siblings</slot0>
</query>
```

Slot Filling

```
CS17_ENG_54d83ec119 per:siblings myrun Doc726:561-577 Bart PER Doc726:434-437 0.7
```

Changes in 2017

- Events and sentiments merged into the KB
- Multiple justifications (K=1 and K=3)
- AP-based scoring

- 1 Introduction
 - Task Variants
 - Changes in 2017
- 2 SF/KB Evaluation
 - Definitions
 - Queries
 - Participants
 - Results
- 3 Conclusion

SF/KB Definitions

Scoring - F1

- Wrong or ineXact is *Spurious*
- Hop 1 filler whose Hop 0 parent filler is Wrong or ineXact, is *Spurious*
- Correct responses are grouped into equivalence classes (EC). At most one response is *Right*; all other *Spurious*
- NAM mention in EC, or NOM mentions and the EC is NOM, then one is *Right*; otherwise, if only NOM mentions in a NAM EC, then one is *Ignored*
- **Reference** = number of single-valued pseudo-slots with a correct response + number of equivalence classes for all list-valued pseudo-slots
- Recall = $\#Right / Reference$
- Precision = $\#Right / (\#Right + \#Spurious)$
- F1 = $2 * Precision * Recall / (Precision + Recall)$
- **Applied only to queries with a known correct answer**

SF/KB Definitions

Scoring - AP

```
=====
QUERY_ID:          CSSF17_ENG_afddc4ed21
LEVEL:             1
AP:                0.3750 = (0.6667 + 1.6667/2)/4
NUM_GROUND_TRUTH: 4
GROUND TRUTH:
  CSSF17_ENG_afddc4ed21:1:1
  CSSF17_ENG_afddc4ed21:1:2
  CSSF17_ENG_afddc4ed21:2:1
  CSSF17_ENG_afddc4ed21:2:2
RANKING:
.....
R.  NODEID                                     CONF.  MAPPED_EC                                     V
-----
1  CSSF17_ENG_afddc4ed21:Entity102:Entity111  0.7396  CSSF17_ENG_afddc4ed21:1:2  0.6667
2  CSSF17_ENG_afddc4ed21:Entity102:Entity110  0.6001  CSSF17_ENG_afddc4ed21:1:1  1.0000
3  CSSF17_ENG_afddc4ed21:Entity103:Entity111  0.4653  -                             0.0000
4  CSSF17_ENG_afddc4ed21:Entity104:Entity110  0.4581  -                             0.0000
5  CSSF17_ENG_afddc4ed21:Entity104:Entity111  0.4513  -                             0.0000
6  CSSF17_ENG_afddc4ed21:Entity103:Entity110  0.4172  -                             0.0000
=====
```

SF/KB Definitions

Aggregates and Projections

P/R/F1

Score Variants	Aggregates Reported	
	Micro-average	Macro-average
SF	Yes	Yes
LDC-MAX	Yes	Yes
LDC-MEAN	No	Yes

AP

Score Variants	Aggregates Reported	
	Micro-average	Macro-average
SF	No	Yes
LDC-MAX	No	No
LDC-MEAN	No	Yes

- SF: consider all entrypoints as a separate query.
- LDC-MAX: Considering the run's best entrypoint per LDC query on the basis of F1 score across both hops.
- LDC-MEAN: Precision, Recall, and F1 for each LDC query is the mean Precision, mean Recall, and mean F1 for all entrypoints for that LDC query.

Micro-averages are computed as:

$$Total_Precision = \frac{Total_Right}{Total_Right + Total_Wrong}$$

$$Total_Recall = \frac{Total_Right}{Total_GT}$$

$$Total_F1 = \frac{2 \times Total_Precision \times Total_Recall}{Total_Precision + Total_Recall}$$

Macro-averages are computed as the mean Precision, mean Recall, and mean F1.

Queries

Slot-types

LDC Queries

	(Queries restricted to) Slot-types			Total
	Slot-filling	Events	Sentiments	
Developed	839	315	238	1392
Pooled	266	187	124	577

SF Queries

	(Queries restricted to) Slot-types			Total
	Slot-filling	Events	Sentiments	
Developed	4063	1528	1150	6741
Pooled	1313	913	602	2828

LDC Queries

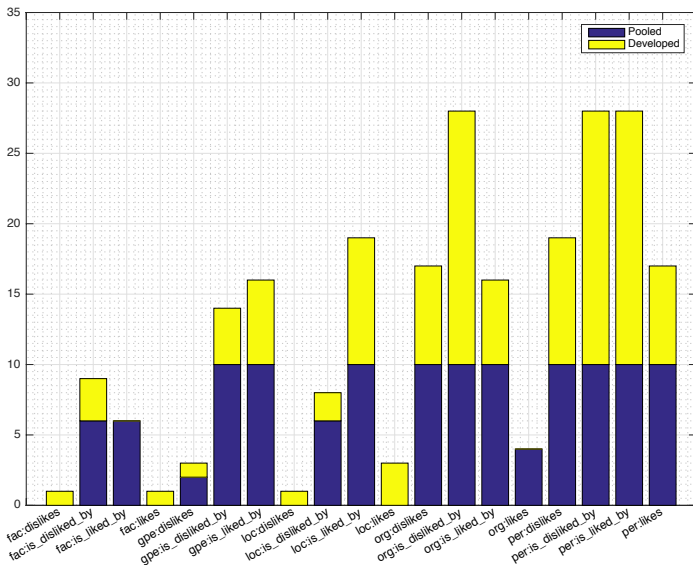
	Language		
	English	Spanish	Chinese
Developed	1300	1198	1112
Pooled	551	509	494

SF Queries

	Language			Total
	English	Spanish	Chinese	
Developed	3282	1870	1589	6741
Pooled	1337	780	711	2828

Queries

LDC Queries - Sentiment Slots



Participants

SF/KB Teams - Submissions with respect to languages

Cold Start 2017

Teams	KB runs					SF runs					Tot.
	ENG	SPA	CMN	XLING	Tot.	ENG	SPA	CMN	XLING	Tot.	
A2KD_Adept	4	-	4	2	10	-	-	-	-	-	10
ICTCAS_OKN	5	-	-	-	5	-	-	-	-	-	5
ISCAS_Sogou	-	-	5	-	5	-	-	-	-	-	5
LDC	1	1	1	1	4	-	-	-	-	-	4
SAFT_ISI	1	1	1	-	3	-	-	-	-	-	3
STANFORD	5	2	4	1	12	4	2	4	1	11	23
TinkerBell	5	5	5	5	20	-	-	-	-	-	20
UNIST_SAIL	-	-	-	-	-	2	-	-	-	2	2
Y_dcd_zju	-	-	-	-	-	3	-	-	-	3	3
dws_uma	-	-	-	-	-	1	-	-	-	1	1
hltcoe	4	-	3	4	11	-	-	-	-	-	11
newbie_mr_01	1	-	-	-	1	-	-	-	-	-	1
Total	26	9	23	13	71	10	2	4	1	17	88

Cold Start 2016

Teams	KB runs					SF runs					Tot.
	ENG	SPA	CMN	XLING	Tot.	ENG	SPA	CMN	XLING	Tot.	
Total	28	9	6	13	56	43	7	7	10	67	123

Participants

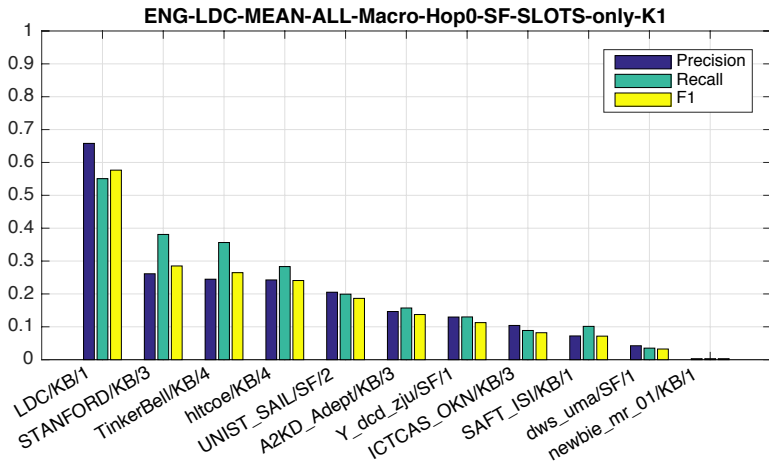
SF/KB Teams - Submissions with respect to slot types

How many submissions had at least one response for a hop-0 query with the given slot type?

Teams	KB runs			SF runs		
	Slot-filling	Events	Sentiments	Slot-filling	Events	Sentiments
A2KD_Adept	10	10	-	-	-	-
ICTCAS_OKN	5	-	2	-	-	-
ISCAS_Sogou	5	5	5	-	-	-
LDC	4	4	4	-	-	-
SAFT_ISI	1	3	3	-	-	-
STANFORD	12	-	-	11	-	-
TinkerBell	20	20	20	-	-	-
UNIST_SAIL	-	-	-	2	-	-
Y_dcd_zju	-	-	-	3	-	-
dws_uma	-	-	-	1	-	-
hltcoe	11	-	-	-	-	-
newbie_mr_01	1	-	-	-	-	-

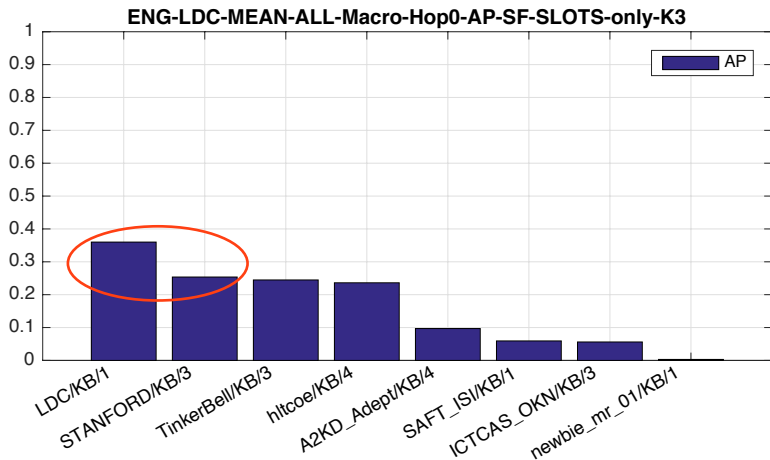
SF/KB Results

K1 - LDC-MEAN-ALL-Macro - Hop 0 - F1 (SF-SLOTS-only) - English



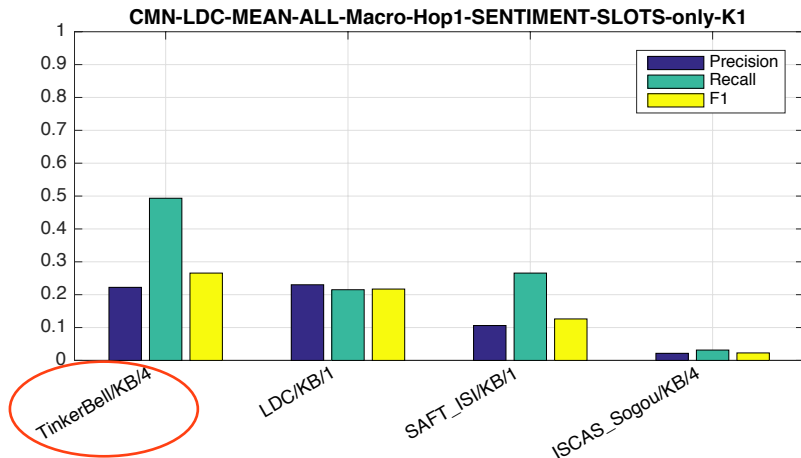
SF/KB Results

K3 - LDC-MEAN-ALL-Macro - Hop 0 - AP (SF-SLOTS-only) - English



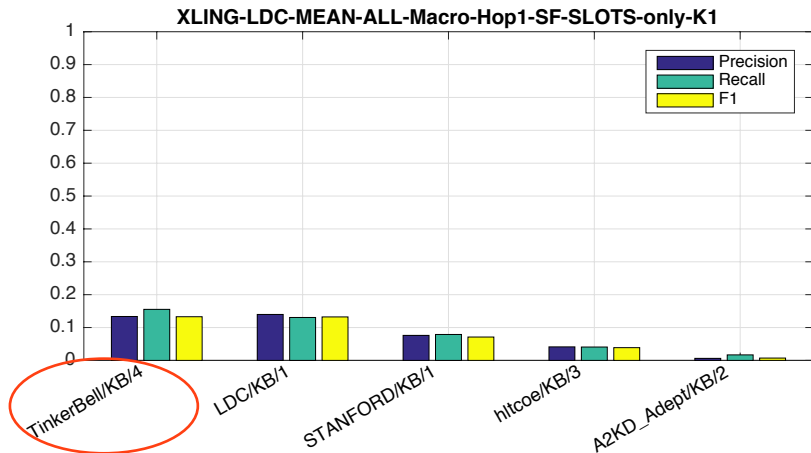
SF/KB Results

K1 - LDC-MEAN-ALL-Macro - Hop 1 - F1 (SENTIMENT-SLOTS-only) - Chinese



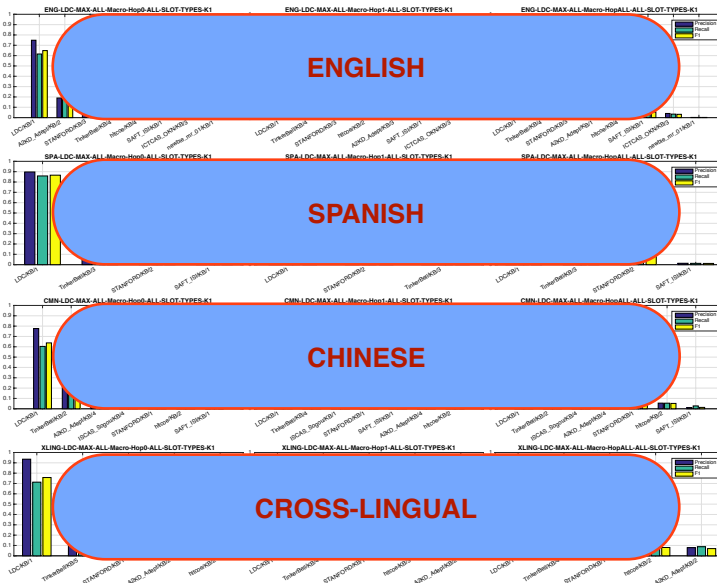
SF/KB Results

K1 - LDC-MEAN-ALL-Macro - Hop 1 - F1 (SF-SLOTS-only) - Cross-lingual



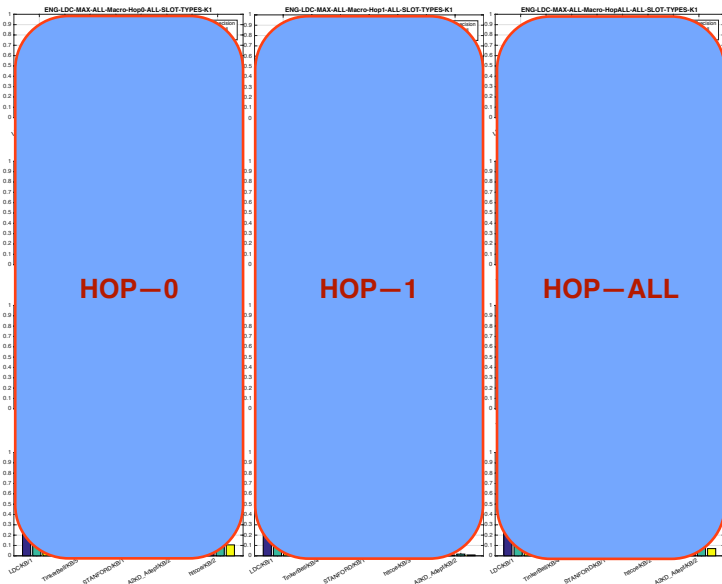
SF/KB Results

Layout



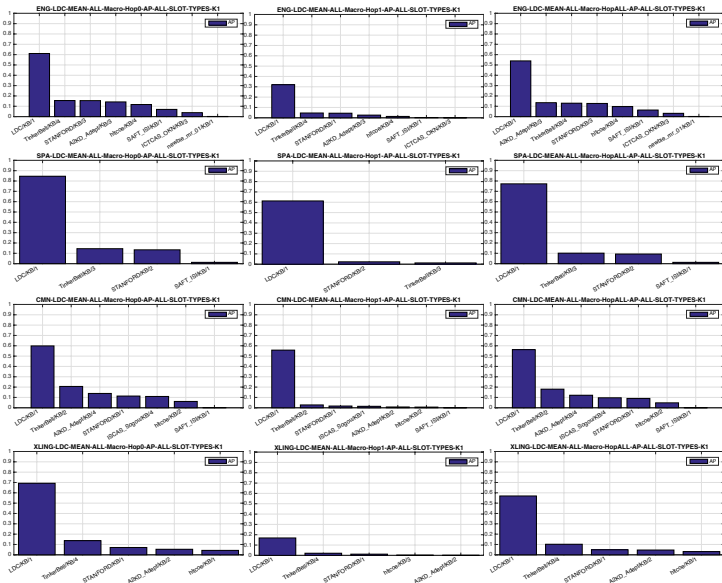
SF/KB Results

Layout



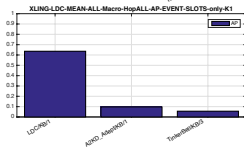
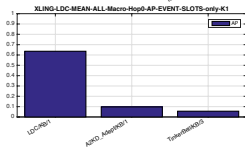
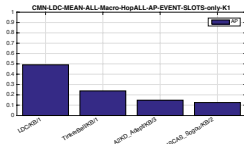
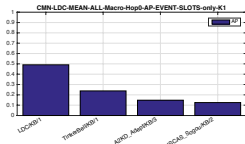
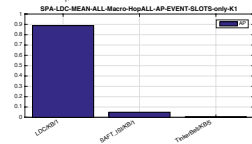
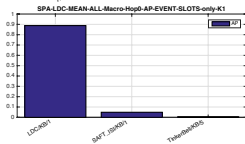
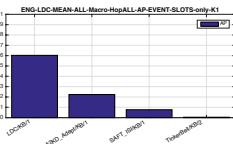
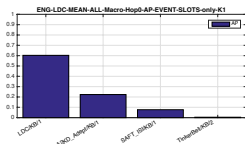
SF/KB Results

K1 - LDC-MEAN-ALL-Macro - AP (ALL-SLOT-TYPES)



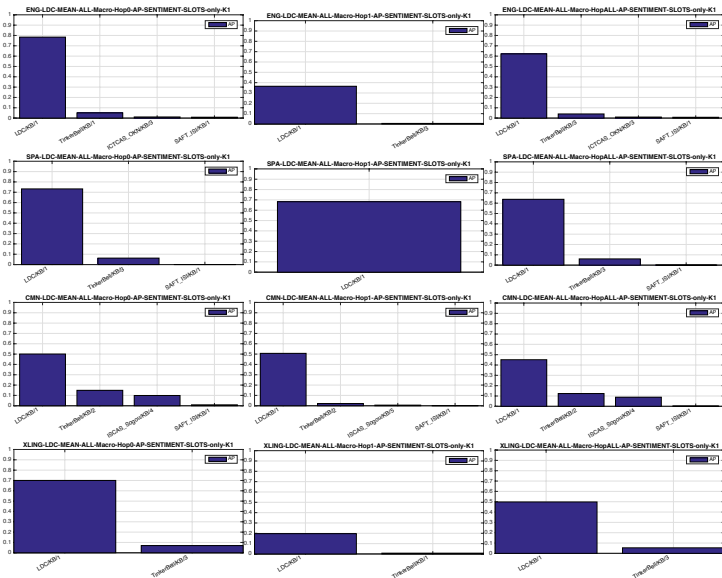
SF/KB Results

K1 - LDC-MEAN-ALL-Macro - AP (EVENT-SLOTS-only)



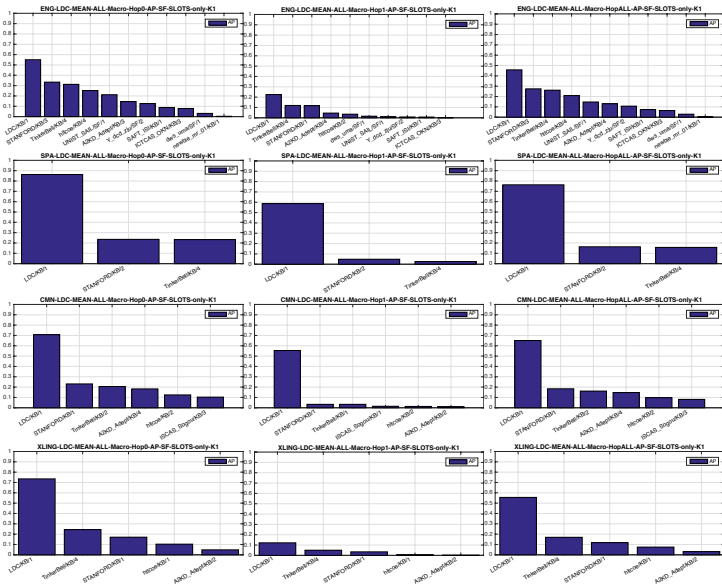
SF/KB Results

K1 - LDC-MEAN-ALL-Macro - AP (SENTIMENT-SLOTS-only)



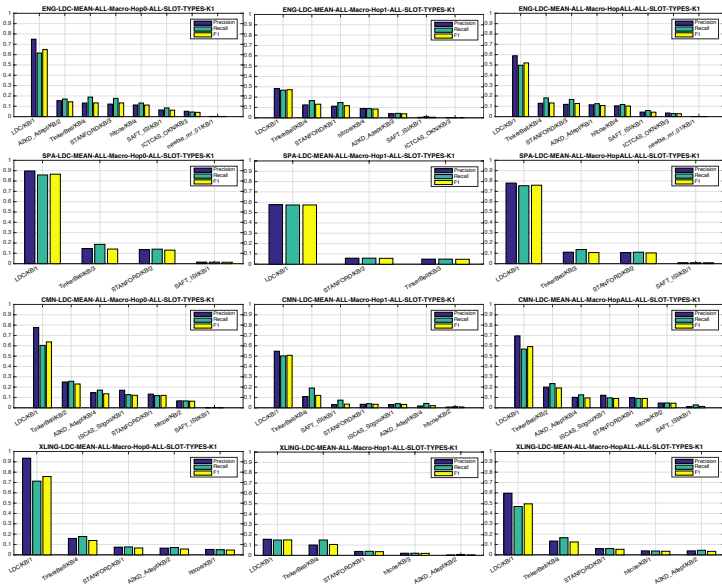
SF/KB Results

K1 - LDC-MEAN-ALL-Macro - AP (SF-SLOTS-only)



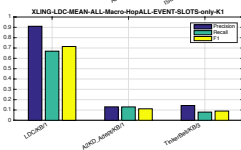
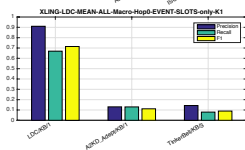
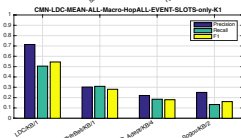
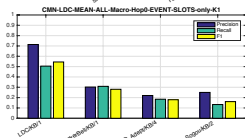
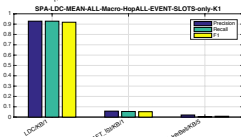
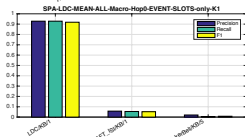
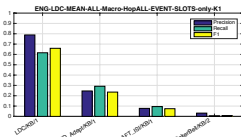
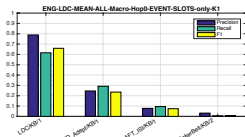
SF/KB Results

K1 - LDC-MEAN-ALL-Macro - F1 (ALL-SLOT-TYPES)



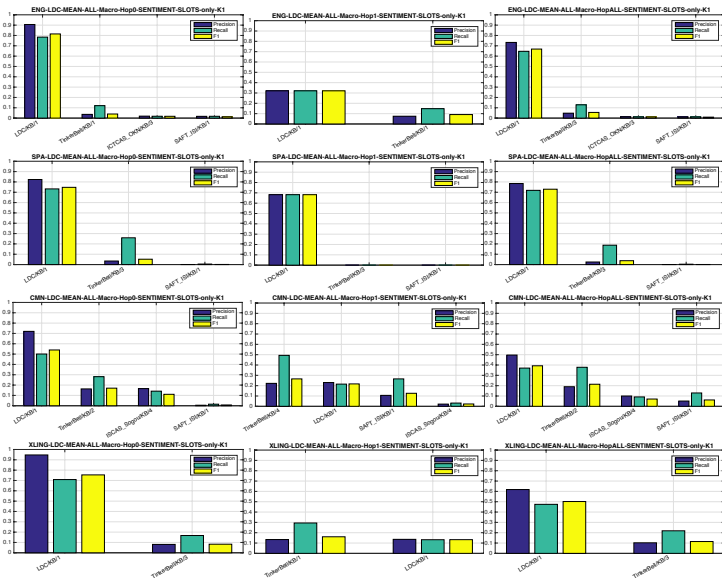
SF/KB Results

K1 - LDC-MEAN-ALL-Macro - F1 (EVENT-SLOTS-only)



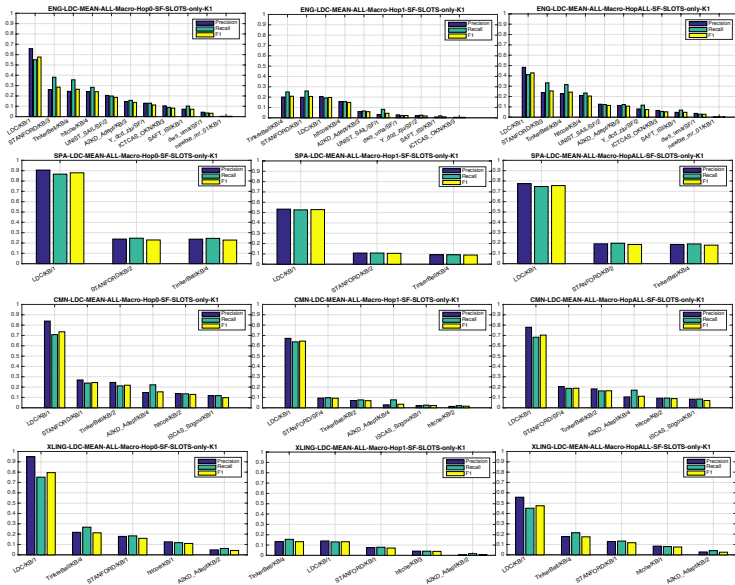
SF/KB Results

K1 - LDC-MEAN-ALL-Macro - F1 (SENTIMENT-SLOTS-only)



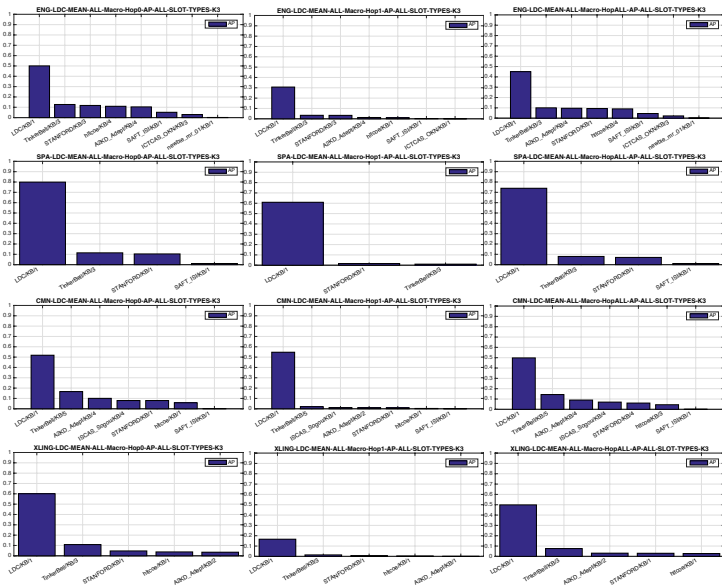
SF/KB Results

K1 - LDC-MEAN-ALL-Macro - F1 (SF-SLOTS-only)



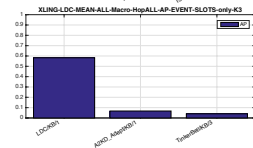
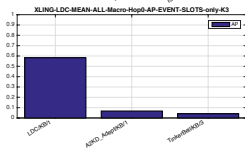
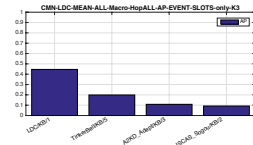
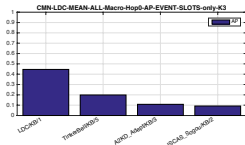
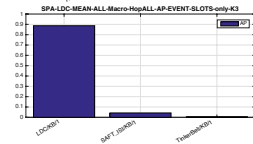
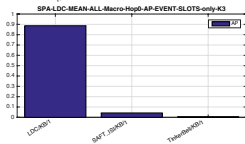
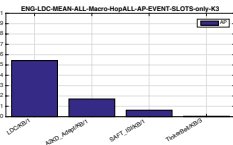
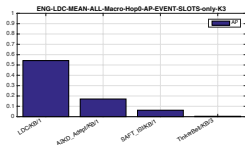
SF/KB Results

K3 - LDC-MEAN-ALL-Macro - AP (ALL-SLOT-TYPES)



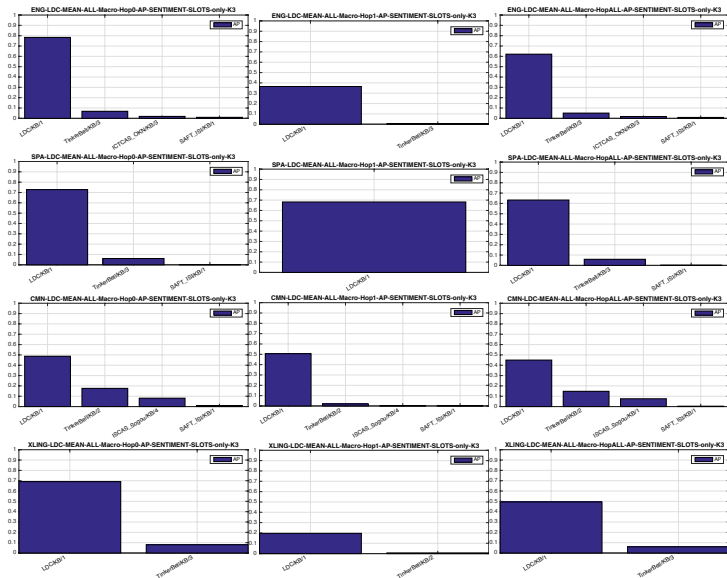
SF/KB Results

K3 - LDC-MEAN-ALL-Macro - AP (EVENT-SLOTS-only)



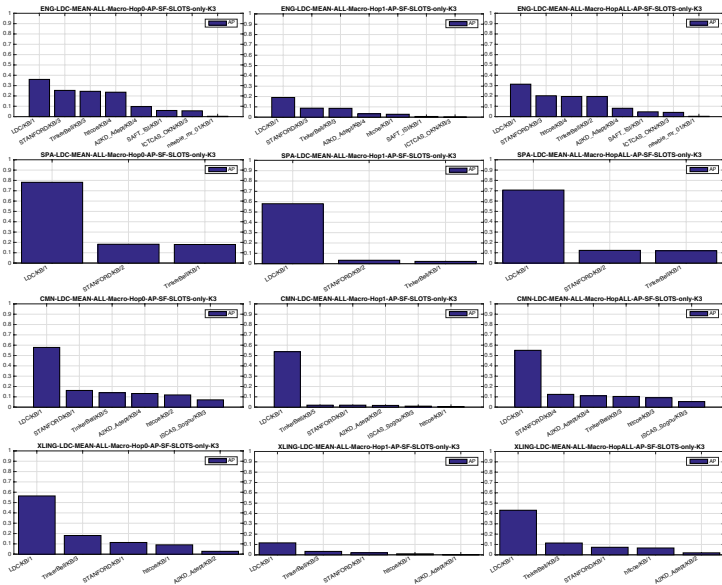
SF/KB Results

K3 - LDC-MEAN-ALL-Macro - AP (SENTIMENT-SLOTS-only)

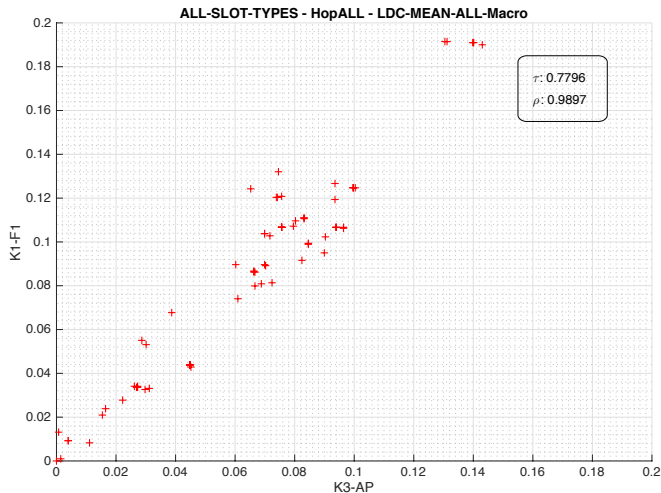


SF/KB Results

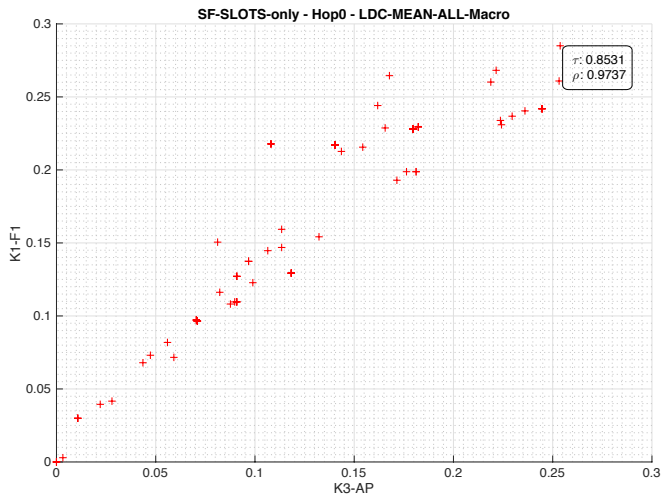
K3 - LDC-MEAN-ALL-Macro - AP (SF-SLOTS-only)



Correlation between AP and F1



Correlation between AP and F1



- 1 Introduction
 - Task Variants
 - Changes in 2017
- 2 SF/KB Evaluation
 - Definitions
 - Queries
 - Participants
 - Results
- 3 Conclusion

Conclusion

- AP scores have near perfect correlation with that of F scores
- Generally automatic systems appear to get about 70% or less performance as compared to manual systems.
- In some cases, we see automatic systems perform better manual system
- Significantly more number of KB submissions as compared to SF submissions
- Much room for improvement on Hop-1 performance
- Much room for improvement on event- and sentiment-based queries