Overview of TAC-KBP2014 Entity Discovery and Linking Tasks

Heng Ji[1], Joel Nothman[2], Ben Hachey[2]

[1] Computer Science Department, Rensselaer Polytechnic Institute
jih@rpi.edu
[2] School of Information Technologies, University of Sydney
{joel.nothman,ben.hachey}@gmail.com

Abstract

In this paper we give an overview of the Entity Discovery and Linking tasks at the Knowledge Base Population track at TAC 2014. In this year we introduced a new end-to-end English entity discovery and linking task which requires a system to take raw texts as input, automatically extract entity mentions, link them to a knowledge base, and cluster NIL mentions. In this paper we provide an overview of the task definition, annotation issues, successful methods and research challenges associated with this new task. This new task has attracted a lot of participants and has intrigued many interesting research problems and potential approaches. We believe it's a promising task to be extended to a tri-lingual setting in KBP2015.

1 Introduction

From 2009 to 2013 the Entity Linking (EL) track at NIST Text Analysis Conference's Knowledge Base Population (TAC-KBP) track aimed to link a given named entity mention from a source document to an existing Knowledge Base (KB). An EL system is also required to cluster mentions for those NIL entities that don't have corresponding KB entries.

Most earlier EL work in the KBP community is usually formulated as a ranking problem, either by (i) Non-collective approaches, which resolve one mention at each time relying on prior popularity, context similarity, and other local features with supervised models (Mihalcea and Csomai, 2007; Milne and Witten, 2008; Han and Sun, 2011; Guo et al., 2013); or (ii) Collective approaches, which disambiguate a set of relevant mentions simultaneously by leveraging the global topical coherence between entities through graph-based approaches (Cucerzan, 2007; Milne and Witten, 2008; Han and Zhao, 2009; Kulkarni et al., 2009; Pennacchiotti and Pantel, 2009; Ferragina and Scaiella, 2010; Fernandez et al., 2010; Radford et al., 2010; Cucerzan, 2011; Guo et al., 2011; Han et al., 2011; Ratinov et al., 2011; Kozareva et al., 2011; Hoffart et al., 2011; Cassidy et al., 2012; Shen et al., 2013; Liu et al., 2013; Huang et al., 2014b). On the other hand, a lot of research has been done in parallel in the Wikification community (Bunescu, 2006) which aims to extract prominent ngrams as concept mentions, and link each concept mention to the KB. One important research direction of the KBP program is "Cold-start", namely we aim to develop an automatic system to construct a KB from scratch. To meed these new needs and challenges, this year we designed a new task called Entity Discovery and Linking (EDL). An EDL system is required to take a source collection of raw documents as input, identify and classify all entity mentions, link each entity mention to the KB, and cluster all NIL mentions. In 2014 we have explored it only for English, but in next year we aim to extend it to the tri-lingual setting with source corpora from three languages: English, Chinese and Spanish.

Compared to the KBP entity linking evaluations in previous years, the main changes and improvement in KBP2014 include:

- Extend English task to Entity Discovery and Linking (full Entity Extraction + Entity Linking + NIL Clustering);

- Add discussion forums to Cross-lingual tracks (some entity morphs naturally exist in Chinese discussion forums);

- Provide basic NLP, full IE, Entity Linking and semantic annotations for some source documents to participants;

- Share some source collections and queries with regular and cold-start slot filling tracks, to investigate the role of EDL in the entire cold-start KBP pipeline.

The rest of this paper is structured as follows. Section 2 describes the definition of each Entity Discovery and Linking task in KBP2014. Section 3 briefly summarizes the participants. Section 4 highlights some annotation efforts. Section 5 and 6 summarize the general architecture of each task's systems and evaluation results, and provide some detailed analysis and discussion. From each participant, we only select the best submission without Web access for comparison. Section 7 discusses the experimental results and sketches our future work.

2 Task Definition and Evaluation Metrics

This section will summarize the Entity Discovery and Linking tasks conducted at KBP 2014. More details regarding data format and scoring software can be found in the KBP 2014 website[1].

2.1 Mono-lingual Entity Discovery and Linking Task

2.1.1 Task Definition

Based on the above motivations, this year we added a new task of Entity Discovery and Linking (EDL) in the mono-lingual English track. The goal is to conduct end-to-end entity extraction, linking and clustering. Given a document collection, an EDL system is required to automatically extract (identify and classify) entity mentions ("queries"), link them to the KB, and cluster NIL mentions (those that don't have corresponding KB entries). Compared to Entity Linking from previous years, an EDL system needs to extract queries automatically. In contrast

to Wikification (Bunescu, 2006; Mihalcea and Csomai, 2007; Ratinov and Roth, 2012), EDL only focuses on three types of entities (Person (PER), Organization (ORG) and Geo-political Entity (GPE, a location with a government)) and requires NIL clustering. In order to evaluate the impact of entity name mention extraction on this new EDL task, we also organized a diagnostic evaluation on English entity linking as defined in KBP2013, with perfect entity name mentions as input.

The input to EDL is a set of raw documents. We selected a subset of the TAC 2014 document collection from multiple genres including newswire, web data, and discussion forum posts, which include high values in terms of both ambiguity and variety and substantial amount of NIL entity mentions. An EDL system is required to automatically generate the following two files.

(1). Mention Query File: An EDL system is required to identify and classify name mentions into person (PER), organization (ORG) or geo-political entity (GPE); and then represent each name mention as a query that consists of a name string, a document ID, and a pair of UTF-8 character offsets indicating the beginning and end locations of the name string in the document. The detailed definition of an entity name mention (a query) is presented in the LDC query development guideline [2]. For example:

```
⟨query id=``EDL14_ENG_0001"⟩
⟨name⟩cairo⟨/name⟩
⟨docid⟩bolt-eng-DF-200-192451-5799099⟨/docid⟩
⟨beg⟩2450⟨/beg⟩
⟨end⟩2454⟨/end⟩
⟨/query⟩
```

(2). Link ID File: Then for each entity mention query, an EDL system should attempt to link it to the given knowledge base (KB). The EDL system is also required to cluster queries referring to the same non-KB (NIL) entities and provide a unique ID for each cluster, in the form of NILxxxx (e.g., "NIL0021"). It should generate a link ID file that consists of the entity type of the query, the ID of the KB entry to which the name refers, or a "NILxxxx" ID if there is no such KB entry, and a confidence value.

---

[1]http://nlp.cs.rpi.edu/kbp/2014/

[2]http://nlp.cs.rpi.edu/kbp/2014/elquery.pdf

Short name	Name in scoring software	Filter	Key	Evaluates	Task
Mention evaluation					
NER	strong_mention_match	NA	*span*	NER	EDL
NERC	strong_typed_mention_match	NA	*span,type*	NER and classification	EDL
Linking evaluation					
NERL	strong_all_match	NA	*span,kbid*	(NER and) linking	EL,EDL
NEL	strong_link_match	*is linked*	*span,kbid*	Link recognition	Diagnostic
NEN	strong_nil_match	*is nil*	*span*	NIL recognition	Diagnostic
Tagging evaluation					
KBIDs	entity_match	*is linked*	*docid,kbid*	Document tagging	Diagnostic
Clustering evaluation					
CEAFm	mention_ceaf	NA	*span*	NER and clustering	EDL
B-Cubed	b_cubed	NA	*span*	Clustering	EL
B-Cubed+	b_cubed_plus	NA	*span,kbid*	Linking and clustering	EL

Table 1: Evaluation measures for entity discovery and linking, each reported as P, R, and F_1. *Span* is shorthand for (*document identifier, begin offset, end offset*). *Type* is PER, ORG or GPE. *Kbid* is the KB identifier or NIL.

2.1.2 Scoring Metrics

Table 1 lists the official evaluation measures for TAC 2014 entity discovery and linking (EDL). It also lists measures for the diagnostic entity linking (EL) task, which are identical to the TAC 2011-13 evaluations. Since systems use gold standard mentions for EL, it isolates linking and clustering performance. The scorer is available at `https://github.com/wikilinks/neleval`.

Set-based metrics Recognizing and linking entity mentions can be seen as a tagging task. Here evaluation treats an annotation as a set of distinct tuples, and calculates precision and recall between gold (G) and system (S) annotations:

$$P = \frac{|G \cap S|}{|S|} \qquad R = \frac{|G \cap S|}{|G|}$$

For all measures P and R are combined as their balanced harmonic mean, $F_1 = \frac{2PR}{P+R}$.

By selecting only a subset of annotated fields to include in a tuple, and by including only those tuples that match some criteria, this metric can be varied to evaluate different aspects of systems (cf. Hachey et al. (2014) which also relates such metric variants to the entity disambiguation literature). As shown in Table 1, NER and NERC metrics evaluate mention detection and classification, while NERL measures linking performance but disregards entity type and NIL clustering. In the EL task where mentions are given, NERL is equivalent to the linking accuracy score reported in previous KBP evaluations.

Results below also refer to other diagnostic measures, including NEL which reports linking (and mention detection) performance, discarding NIL annotations; NEN reports the performance of NIL annotations alone. KBIDs considers the set of KB entities extracted per document, disregarding mention spans and discarding NILs. This measure, elsewhere called *bag-of-titles evaluation*, does not penalize boundary errors in mention detection, while also being a meaningful task metric for document indexing applications of named entity disambiguation.

Clustering metrics Alternatively, entity linking is understood as a cross-document coreference task, in which the set of tuples is partitioned by the assigned entity ID (for KB and NIL entities), and a coreference evaluation metric is applied.

Where previous KBP evaluations employed B-Cubed (Bagga and Baldwin, 1998), 2014 is the first year to apply CEAF (Luo, 2005). B-Cubed assesses the proportion of each tuple's cluster that is shared between gold and system clusterings, while CEAF calculates the optimal one-to-one alignment between system and gold clusters based on a provided similarity metric, and returns the sum of aligned scores relative to aligning each cluster with itself. In the Mention CEAF (CEAFm) variant used here, cluster similarity is simply measured as the number of overlapping tuples.

Again, variants may be introduced by selecting a subset of fields or filtering tuples. For the present work, only the mention's span is considered, except in B-Cubed+ which treats mention as *span,kbid* tuples, only awarding mentions that are clustered together if their KB link is correct.

We now define the clustering metrics formally. If we let $G_i \in \mathcal{G}$ describe the gold partitioning, and $S_i \in \mathcal{S}$ for the system, we calculate the

maximum score bijection m:

$$m = \arg\max_m \sum_{i=1}^{|\mathcal{G}|} |G_i \cap S_{m(i)}|$$
$$\text{s.t. } m(i) = m(j) \iff i = j$$

Then we can define `CEAFm`, and provide `B-Cubed` for comparison:

$$P_{\text{CEAFm}} = \frac{\sum_{i=1}^{|\mathcal{G}|} |G_i \cap S_{m(i)}|}{\sum_{i=1}^{|\mathcal{S}|} |S_i|}$$

$$R_{\text{CEAFm}} = \frac{\sum_{i=1}^{|\mathcal{G}|} |G_i \cap S_{m(i)}|}{\sum_{i=1}^{|\mathcal{G}|} |G_i|}$$

$$P_{\text{B-Cubed}} = \frac{\sum_{i=1}^{|\mathcal{G}|} \sum_{j=1}^{|\mathcal{S}|} \frac{|G_i \cap S_j|^2}{|S_j|}}{\sum_i |S_i|}$$

$$R_{\text{B-Cubed}} = \frac{\sum_{i=1}^{|\mathcal{G}|} \sum_{j=1}^{|\mathcal{S}|} \frac{|G_i \cap S_j|^2}{|G_i|}}{\sum_i |G_i|}$$

Like the set-based metrics, `B-Cubed` and `CEAFm` report the number of mentions that are correct by some definition. Hence these metrics are likely to be correlated, and `B-Cubed+` is by definition bounded from above by both `B-Cubed` and `NERL`. CEAF prefers systems that return the correct quantity of clusters, and splitting an entity into multiple clusters mean some of those clusters will be awarded no score. Thus a system that incorrectly splits *Abu Mazen* and *Mahmoud Abbas* into different entities will be assigned no score for the smaller of those clusters (both precision and recall error). In `B-Cubed`, the same system is awarded for correctly predicting that multiple mentions were coreferent with each other, despite the entity being split.

Confidence intervals We calculate $c\%$ confidence intervals for set-based metrics by bootstrap resampling documents from the corpus, calculating these pseudo-systems' scores, and determining their values at the $\frac{100-c}{2}$th and $\frac{100+c}{2}$th percentiles of 2500 bootstrap resamples. This procedure assumes that the system annotates documents independently, and intervals are not reliable where systems use global clustering information in their set-based output (i.e. beyond NIL cluster assignment). For similar reasons, we do not calculate confidence intervals for clustering metrics.

2.2 Cross-lingual Chinese/Spanish to English Entity Linking

The cross-lingual entity linking tasks follow the monolingual entity linking in previous years (Ji et al., 2011) in which the entity mention queries are given; the steps are: (1) link non-NIL queries to English KB entries; and (2) cluster NIL queries.

3 Participants Overview

Table 2 summarizes the participants for English EDL task. In total 20 teams submitted 75 runs for the full task and 17 teams submitted 55 runs for the diagnostic task (Entity Linking with perfect mentions as input).

4 Data Annotation and Resources

The details of the data annotation for KBP2014 are presented in a separate paper by the Linguistic Data Consortium (Ellis et al., 2014). The new EDL task has also introduced many new challenges to the annotation guideline, especially on defining entity mentions in the new setting of cold-start KBP. We released a new KBP entity mention definition guideline [3]. There still remain some annotation errors on both entity mention extraction and linking. We will continue refining the annotation guidelines and conduct a systematic correction on the annotation errors.

In addition, we devoted a lot of time at collecting related publications and tutorials [4], resources and softwares [5] to lower down the entry cost for EDL.

5 Mono-lingual Entity Linking

5.1 Approach Overview

5.1.1 General Architecture

A typical KBP2014 mono-lingual EDL system architecture is summarized in Figure 1. It includes six steps: (1) entity mention extraction - identify and classify entity mentions ("queries") from the source documents; (2) query expansion - expand the query into a richer set of forms using Wikipedia structure mining or coreference resolution in the background document; (3) candidate generation - finding all possible KB entries that a query might link to; (4) candidate ranking - rank the probabilities of all candidates

---

[3]http://nlp.cs.rpi.edu/kbp/2014/elquery.pdf
[4]http://nlp.cs.rpi.edu/kbp/2014/elreading.html
[5]http://nlp.cs.rpi.edu/kbp/2014/tools.html

Team Name	Organization	Full	Diagnostic
BUPT_PRIS	Beijing university of posts and telecommunications	5	4
BUPTTeam	Beijing University of Post and Telecommunication	4	
CCNU_NLP	Central China Normal University	1	
CohenCMU	Carnegie Mellon University	5	3
CSFG	Centre for Structural and Functional Genomics, Concordia University	5	5
HITS	Heidelberg Institute for Theoretical Studies	4	2
IBM	International Business Machines Corporation	2	
ICTCAS_OKN	Institute of Computing Technology, Chinese Academy of Sciences	3	2
insight	Insight Centre for Data Analytics at National University of Ireland, Galway	4	
lcc	Language Computer Corporation	4	5
lkd	University of Economics, Prague	3	
MSIIPL_THU	Department of Electronic Engineering, Tsinghua University	5	5
NYU	New York University	4	5
OSU	Oregon State University	5	
RPI_BLENDER	Rensselaer Polytechnic Institute	4	4
SIEL_TAC	International Institute of Information Techonology, Hyderabad	1	
Tohoku_NL	Graduate School of Information Sciences,Tohoku University	2	
UBC	University of the Basque Country	5	4
UI_CCG	Cognitive Computation Group	4	1
vera	Beijing University of Posts and Telecommunications	5	5
Bobwilson	University of Sydney		1
Compreno	ABBYY		1
MS_MLI	Microsoft Research		2
NTUA_ILSP	National Technical University of Athens		4
TALP_UPC	Universitat PolitÄ¨cnica de Catalunya (UPC)		2

Table 2: The Number of Runs Submitted by KBP2014 English Entity Discovery and Linking Participants

using non-collective or collective approaches; the linking decision (knowledge from the KB) can be used as feedback to refine the entity mention extraction results from step (1); (4) NIL detection and clustering - detect the NILs which got low confidence at matching the top KB entries from step (4), and group the NIL queries into clusters.

Most the ranking algorithms are inherited from previous years, except the novel Progamming with Personalized PageRank algorithm developed by the CohenCMU team (Mazaitis et al., 2014). A nice summary of the state-of-the-art ranking features can be found in Tohoku_NL team's system description paper (Zhou et al., 2014). In the following subsections we will highlight the new and effective techniques used in entity linking.

5.2 Evaluation Results

5.2.1 Overall Performance

The linking and clustering results of mono-lingual EDL are summarized in Figures 2 and 3 respectively.[6]

Although CEAFm does not explicitly take link target into account, its results closely mirror NERL, with some variation: for example, BUPTTeam1 ranks much better under CEAFm than NERL. The confidence intervals (CIs) shown in Figure 2 highlight two main cohorts of high-achieving systems: those with performance around or above 60% NERL F_1 may be significantly different from those below. Only other lcc systems and RPI_BLENDER5 fall in the NERL 90% CI for the top system (lcc20142).

5.2.2 Impact of Mention Extraction

A major change from previous KBP evaluations is considering the end-to-end task in which systems are required to identify, link and cluster mentions. Mention extraction is very important to good NERL and CEAFm performance, as shown by the NER performance in Figure 4.

The KBIDs measure results show that a few systems (particularly UBC5, IBM1) are able to predict the set of correct KB targets for a document but are let down on the span-sensitive metrics by low NER performance. This measure

---

[6]In most figures, we only show the performance of the top system per team on the primary task score (CEAFm for EDL and B-Cubed+ for EL), and ordered according to that score.
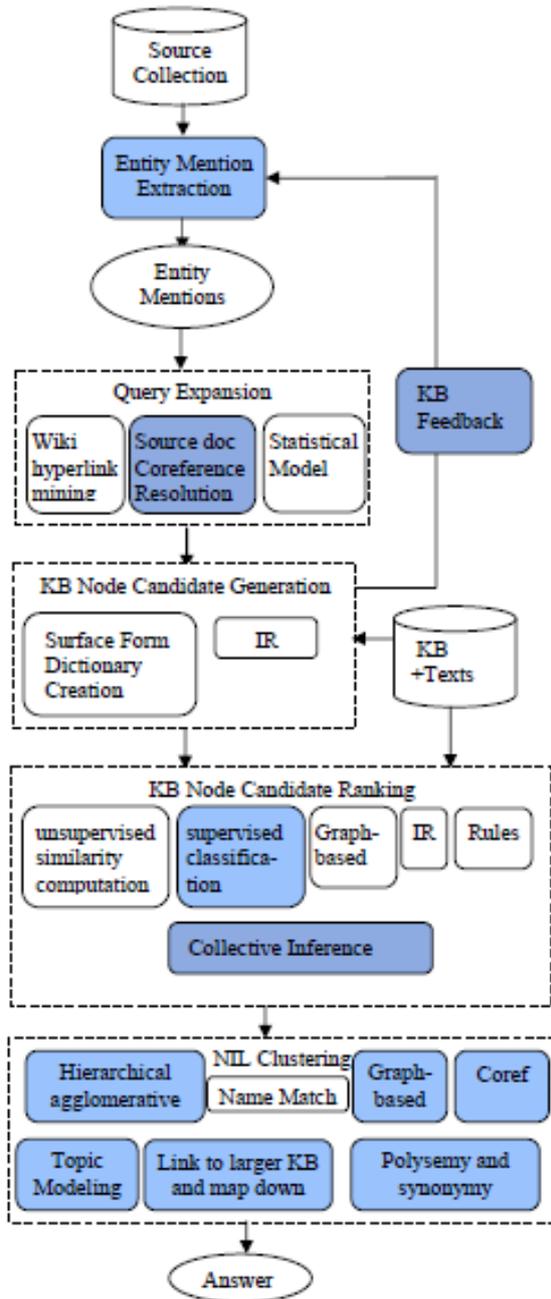
Figure 1: General Mono-lingual Entity Linking System Architecture

Figure 2: Linking performance (`NERL`) of top 14 monolingual EDL systems (`CEAFm` $F_1 > 50\%$) with 90% confidence intervals

Figure 3: Clustering performance (`CEAFm`) of top 14 monolingual EDL systems (`CEAFm` $F_1 > 50\%$)

Figure 4: EDL F_1 measures related to mention extraction

Figure 5: Performance of top 15 Diagnostic EDL (perfect mention as input) systems (`B-Cubed+` $F_1 > 50\%$) under various measures [8]

is also highly correlated with `NERL`/`NER`, the ratio of linking to mention extraction performance, which should isolate the difficulty of linking given noisy system mentions. By this ratio, state-of-the-art linking performance lies between 80 and 90%, with some systems such as IBM1 achieving a very high score by this ratio, while let down by a low `NER` performance. This agrees with the `NERL` results of the diagnostic EL evaluation – which provides systems with gold-standard mentions – shown in Figure 5.
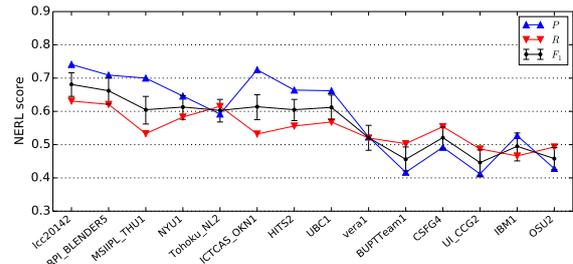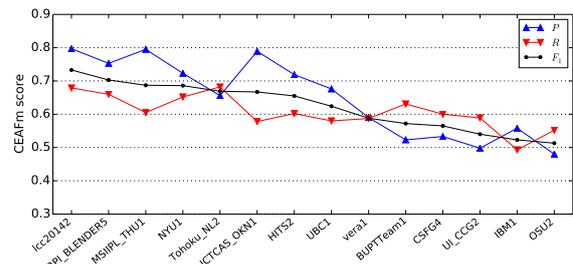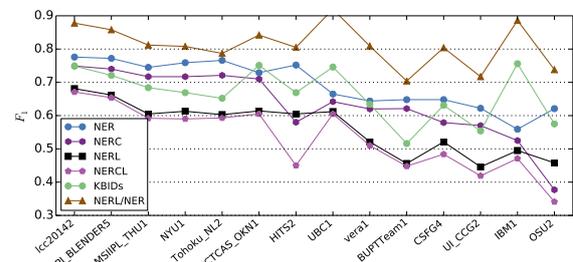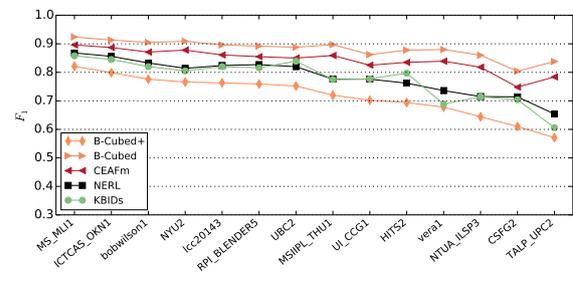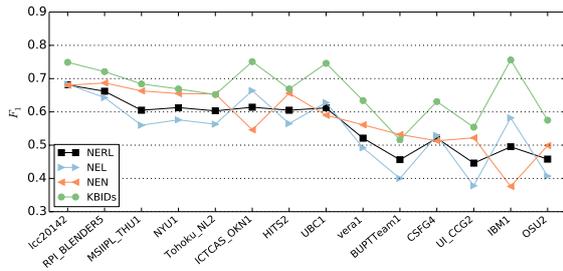
Figure 6: EDL F_1 measures comparing NIL and non-NIL performance

Note also that `KBIDs` corresponds closely with `NERL` in the EL diagnostic task, which suggests that the disparity between `KBIDs` and `NERL` in EDL pertains to `NER` performance rather than in-document coreference.

Comparing `NEL` (linked only) and `NEN` (NIL only) in Figure 6 shows that the systems that perform particularly well on `KBIDs` also handle KB mentions better than NILs. Conversely, the MSIIPL_THU systems rank well by `CEAFm` with a system that handles NILs better than KB mentions.

As well as identifying mentions, systems were required to assign each entity a type, which is useful for downstream KBP tasks such as slot filling. Identifying mentions of entities other than the sought types also leads to precision errors in all other measures. Thus teams with good `NER` also have good entity types (`NERC`), HITS being the notable exception. Performance reduces as little as 3% F_1 due to type detection. Note that top `NERC` performance is only 74.9%, which is much lower than state-of-the-art named tagging performance (around 89% F-score) on standard datasets such as ACE or CONLL (e.g., Ratinov and Roth (2009)). Other tasks incorporate a different set of entity types, such as the catch-all "MISC" type and non-GPE locations; their loss may make the task more difficult. However, systems did not tend to train their recognisers directly on the in-domain corpus, and the lower performance appears consistent with out-of-domain results for `NERC` (Nothman et al., 2013).

5.2.3 Entity Types and Textual Genres

In Figures 7 and 8 we plot, respectively, the EDL and EL `NERL` performance when subsets of the queries are considered, breaking down the collection by entity type and textual genre. For
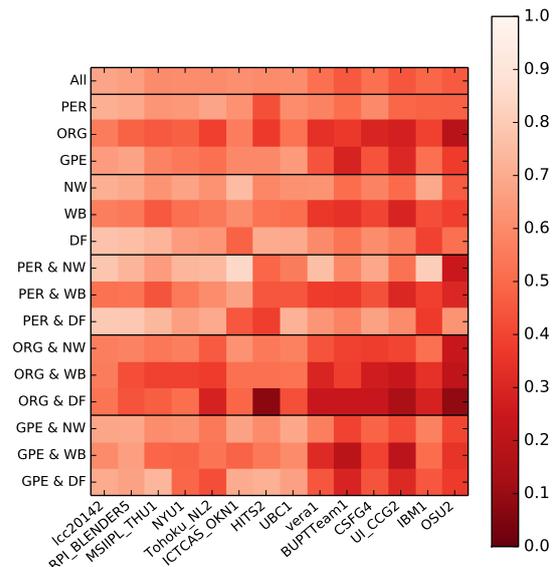

Figure 7: EDL `NERL` F_1 when selecting a subset of annotations by entity type and text genre. PER = person, ORG = organization, GPE = geopolitical entity; NW = newswire, WB = web log, DF = discussion forum.
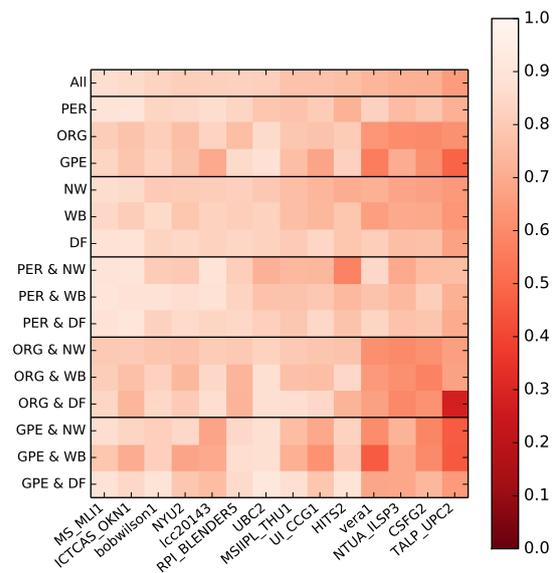

Figure 8: EL `NERL` F_1 when selecting a subset of annotations by entity type and text genre. See key in Figure 7.
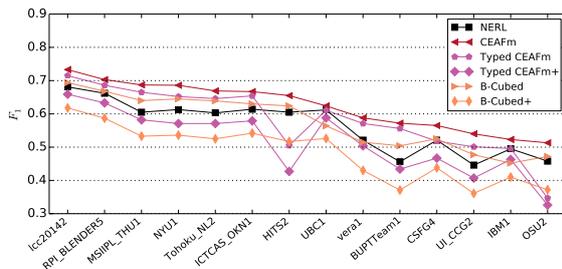
Figure 9: EDL F_1 measures related to clustering

EDL, the gold standard annotations are filtered by gold standard entity type and the system annotations by predicted entity type. `CEAFm` results show similar trends.

Performance is generally much higher on person entities than the other types. GPE clearly places second in EDL, perhaps reflecting the difficulty of `NER` for diverse organization names, particularly out of the news genre. But the ranking of GPE and ORG is more equivocal in EL; UBC and HITS perform markedly better on GPE and ORG than PER.

In both task variants, discussion forum text proved easiest to process, at least for PER and GPE extraction, followed by newswire and web logs. The variation in mention detection performance on PERs in forums derives from the need to identify poster names, not merely names in running text. Systems with low performance here – ITCAS_OKN1, HITS2 and IBM1 – appear not to have specifically targeted these mentions. Recognition of person names in web logs appears particularly difficult relative to other genres, while GPEs are difficult to link (even with gold mentions) in web logs.

5.2.4 Clustering Performance

The newly adopted `CEAFm` measure provides a ranking that is very similar to that of `B-Cubed` or `B-Cubed+`, as shown in Figure 9. This is particularly true across the top six teams, with more variation in the lower rankings. The differences are more pronounced when comparing to `B-Cubed+`, since it is a combined measure of linking and clustering performance.

Notably, the EL task reports higher `B-Cubed` than `CEAFm` scores in all cases (Figure 5), while EDL shows the reverse trend (Figure 9). This highlights the sensitivity of `B-Cubed` to mention extraction errors, as discussed in Cai and Strube

(2010).

5.3 What's New and What Works

5.3.1 EDL Milestones

Overall Entity Linking and Wikification research is a rapidly growing and thriving area across various research communities including Natural Language Processing, Data Mining and Semantic Web, demonstrated by the 130 representative papers published during 2006-2014 [9]. We listed some milestones in the following.

- 2006: The first definition of Wikification task is proposed by (Bunescu, 2006).

- 2009: TAC-KBP Entity Linking was launched (McNamee and Dang, 2009).

- 2008-2012: Supervised learning-to-rank with diverse levels of features such as entity profiling, various popularity and similarity measures were developed (Chen and Ji, 2011; Ratinov et al., 2011; Zheng et al., 2010; Dredze et al., 2010; Anastacio et al., 2011).

- 2008-2013: Collective Inference, Coherence measures were developed (Milne and Witten, 2008; Kulkarni et al., 2009; Ratinov et al., 2011; Chen and Ji, 2011; Ceccarelli et al., 2013; Cheng and Roth, 2013).

- 2012: Various applications(e.g., Knowledge Acquisition (via grounding), Coreference resolution (Ratinov and Roth, 2012) and Document classification (Vitale et al., 2012; Song and Roth, 2014; Gao et al., 2014).

- 2014: TAC-KBP Entity Discovery and Linking (end-to-end name tagging, cross-document entity clustering, entity linking).

- 2012-2014: Many different versions of international evaluations were inspired from TAC-KBP.

5.3.2 Joint Extraction and Linking

Some recent work (Sil and Yates, 2013; Meij et al., 2012; Guo et al., 2013; Huang et al., 2014b) proved that mention extraction and mention linking can mutually enhance each other. Inspired by the these successes, many teams including IBM (Sil and Florian, 2014), MSIIPL_THU (Zhao

---

[9] http://nlp.cs.rpi.edu/kbp/2014/elreading.html

et al., 2014), SemLinker (Meurs et al., 2014), UBC (Barrena et al., 2014) and RPI (Hong et al., 2014) used the properties in external KBs such as DBPedia as feedback to refine the identification and classification of name mentions. Using this approach, RPI system successfully corrected 11.26% wrong mentions. The HITS team (Judea et al., 2014) proposed a joint approach that simultaneously solves extraction, linking and clustering using Markov Logic Networks. For example, in the following sentence "*Bosch will provide the rear axle.*", linking "*Bosch*" to "*Robert Bosch Tool Corporation*" based on context "*rear axle*" can help us type it as an organization. Similarly, it's relatively easy to link "*San Antoni*" in the following sentence: "*Parker was 15 for 21 from the field, putting up a season high while scoring nine of San Antonio's final 10 points in regulation.*" to "*San Antonio Spurs*" and type it as an organization instead of GPE.

5.3.3 Task-specific and Genre-specific Mention Extraction

The new definition of entity mentions in the KBP setting and new genres require us to do some effective adaptation of traditional name tagging approaches. For example, 4% entity mentions included nested mentions. Posters in discussion forum should be extracted. Due to time limit, several teams including HITS (Judea et al., 2014), LCC (Monahan et al., 2014), MSIIPL_THU (Zhao et al., 2014), NYU (Nguyen et al., 2014) and RPI (Hong et al., 2014) developed heuristic rules to improve name tagging. In the future, we expect that more effective genre adaptation approach can be developed to further improve the performance.

5.4 Better Meaning Representation

Addressing many linking challenges requires acquiring and better representing the deeper, richer and more discriminative semantic knowledge of each entity mention and its context, beyond the semantic attributes from typical information extraction, dependency parsing and semantic role labeling techniques. For example, in

> "*Local **OWS** activists were part of this **protest***",

the candidate entities for "*OWS*" include "*Order of World Scouts*", "*Occupy Wall Street*", "*Oily Water Separator*", "*Overhead Weapon Station*",

"*Open Window School*" and "*Open Geospatial Consortium*". The knowledge that "*OWS*" is associated with a "*protest*" event is needed to correctly link it to "*Occupy Wall Street*". Similarly, in the following discussion forum posts:

> "*It was a pool **report** typo. Here is exact **Rhodes quote**: "this is not gonna be a couple of weeks. It will be a period of days." At a **WH briefing** here in Santiago, **NSA** spox **Rhodes** came with a litany of pushback on idea **WH** didn't consult with **Congress**. **Rhodes singled** out a **Senate** resolution that passed on March 1st which denounced Khaddafy's atrocities. **WH says UN** rez incorporates it.*",

in order to link "*Rhodes*" to the speech writer "*Ben Rhodes*" in the KB, we rely on knowledge that a "*speech writer*" usually initiates: "*quote*", "*report*", "*briefing*" and "*singled out*", and the organizations a "*speech writer*" usually interacts with: "*WH*", "*NSA*", "*Congress*", "*Senate*" and "*UN*".

RPI system (Zheng et al., 2014) exploited the Abstract Meaning Representation (AMR) (Banarescu et al., 2013) and an AMR parser (Flanigan et al., 2014) to discover and represent rich knowledge from the source texts. They proved that using AMR a simple unsupervised collective inference method without using any labeled EL data can outperform other representations such as semantic role labeling and all state-of-the-art unsupervised linking.

5.5 Select Collaborators from Rich Context

Many teams including lkd (Dojchinovski et al., 2014), NYU (Nguyen et al., 2014) and RPI (Hong et al., 2014) exploited the rich properties and structures in DBPedia for collective inference. The basic intuition is that the candidate entity and its collaborators decided by the mention's collaborators in the source text should be strongly connected in the KB. Collective inference is particularly effective to disambiguate entities with common names in discussion forum posts. For example, many countries can have a "*Supreme Court*" or "*LDP*"; "*Newcastle University*" can be located in UK or Australia; and many person entities share the same common names such as "*Albert*"; etc..

However, there might be many entity mentions in the context of a target entity

mention that could potentially be leveraged for disambiguation. Various coherence measures were introduced in recent research to choose the "collaborators" (related mentions), such as collaborative learning (Chen and Ji, 2011), ensemble ranking (Pennacchiotti and Pantel, 2009; Kozareva et al., 2011), co-occurred concept mentions (McNamee et al., 2011; Ratinov et al., 2011), topic modeling (Cassidy et al., 2012), relation extraction (Cheng and Roth, 2013), coreference (Huang et al., 2014a), semantic relatedness based meta-paths (Huang et al., 2014a) and social networks (Cassidy et al., 2012; Huang et al., 2014a). RPI system (Zheng et al., 2014) exploited the semantic relation links among concepts from AMR to select collaborators.

Table 3 presents the various types of entity contexts that may help disambiguate entities. In addition, some global context such as document creation time will be helpful for entity disambiguation.

5.6 Graph-based NIL Entity Clustering

The CUNY-BLENDER KBP2012 Entity Linking system (Tamang et al., 2012) explored more than 40 clustering algorithms and found that advanced graph-based clustering algorithms did not significantly our-perform single baseline "All-in-one" clustering algorithm on the overall queries (except the most difficult ones). However, this year it's very encouraging that LCC (Monahan et al., 2014) proved that graph partition based algorithm achieved significant gains.

5.7 Remaining Challenges

Regardless of the progress that we have achieved with EDL, many challenges still remain. We highlight some outstanding ones as follows.

5.7.1 Mention Extraction is not a Solved Problem

The best name mention extraction F-score in the EDL task is only about 75%, which is much lower than the 89% F-score reported in the literature (Ratinov and Roth, 2009; Li et al., 2012). Table refnamemilestone lists some previous milestones for name tagging.

Compared to the first name tagging paper in 1966, we made good progress on developing machine learning algorithms and incorporating a few more features. There hasn't been a lot of active work in the field after ACE because we we

tend to believe it's a solved problem. Dojchinovski et al. (2014) provided some detailed analysis on mention extraction challenges in EDL, especially on the confusion between ORG and GEP (e.g., between sports teams and host cities). In addition, we summarized the new challenges as follows.

- Our name taggers are getting old. Almost all of them were trained from 2003 news but now tested on 2012 news.

- Need effective genre adaption techniques. Table 5 shows the performance of a state-of-the-art name tagger (Li et al., 2012) trained and tested from various genre combinations. We can see that a tagger trained from newswire can obtain 89.3% F-score on newswire but only 56.2% on broadcast news.

- Need to address the new definitions of name mention in the cold-start KBP setting by following the "extraction for linking" philosophy.

- Need to re-visit some old unsolved problems. For example, it remains difficult to identify long nested organization names such as "Asian Pulp and Paper Joint Stock Company , Lt. of Singapore". We should also continue to explore external knowledge sources such as Word clustering, Lexical Knowledge Discovery (Ratinov and Roth, 2009; Ji and Lin, 2009) and more feedback from entity linking, coreference, relation and event extraction to improve the quality of name tagging. For example, without knowing that "FAW" referring to "First Automotive Works" in the following sentence "FAW has also utilized the capital market to directly finance,. . . ", it's very challenging to classify "FAW" as an organization name.

5.7.2 Linking Challenge: Normalization

Currently there is no normalization schema to easily align the property types across multiple KBs with different naming functions (e.g., *extinction* vs. *date of dissolved* for an organization) and granularities (e.g., *birthYear* vs. *date of birth* for an organization). It's also challenging to automatically map the relation types among entities in source documents and the property types in KBs. For example, when a person

Source	Knowledge Base
Social Relation No matter what, he never should have given **[Michael Jackson]c** that propofol. He seems to think a "proper" court would have let **[Murray]m** go free.	**relation** **[Conrad Murray]e** was charged with involuntary manslaughter for causing **[Michael Jackson]c**'s death on June 25, 2009, from a massive overdose of the general anesthetic propofol.
Family **[Mubarak]m**, the wife of deposed Egyptian President **[Hosni Mubarak]c**, . . .	**relation** **[Suzanne Mubarak]e** (born 28 February 1941) is the wife of former Egyptian President **[Hosni Mubarak]c** and was the First Lady of Egypt during her husband's presidential tenure from 14 October 1981 to 11 February 2011.
Employment Hundreds of protesters from various groups converged on the state capitol in Topeka, **[Kansas]c** today. . . Second, I have a really hard time believing that there were any ACTUAL "explosives" since the news story they link to talks about one guy getting arrested for THREATENING Governor **[Brownback]m**.	**relation** **[Sam Bronwnback]e** was elected Governor of **[Kansas]c** in 2010 and took office in January 2011.
Affiliation During the 19th century of the **[United States]c**, the **[Republican Party]m** stood against the extension of slavery to the country's new territories and, ultimately, for slavery's complete abolition.	**relation** The **[Republican Party]e**, also commonly called the GOP (abbreviation for Grand Old Party), is one of the two major contemporary political parties in the **[United States]c**.
Part-whole AT&T coverage in **[GA]c** is good along the interstates and in the major cities like Atlanta, Athens, **[Rome]m**, Roswell, and Albany.	**relation** At the 2010 census, **[Rome]e** had a total population of 36,303, and is the largest city in Northwest **[Georgia]c** and the 19th largest city in the state.
Start-Position Event Going into the big **[Super Tuesday]c**, **[Romney]m** had won the most votes, states and delegates, **[Santorum]c** had won some contests and was second, **[Gingrich]c** had won only one contest.	**relation** The **[Super Tuesday]c** primaries took place on March 6. **[Mitt Romney]m** carried six states, **[Rick Santorum]c** carried three, and **[Newt Gingrich]c** won only in his home state of Georgia.

Table 3: Various Types of Entity Context

Training Domain	Test Domain	Prec	Recall	F_1
nwire	nwire	89.6	89.1	89.3
	bn	92.5	40.3	56.2
	bc	87.3	84.4	85.8
	cts	71.4	85.6	77.8
	wl	77.2	73.2	75.2
	un	77.7	59.4	67.3
bn	nwire	84.8	77.7	81.2
	bn	89.6	85.1	87.3
	bc	87.1	81.6	84.3
	cts	85.2	83.1	84.2
	wl	75.3	69.8	72.5
	un	75.5	57.8	65.5
bc	bc	88.4	82.5	85.4
cts	cts	91.9	77.9	84.3
wl	wl	75.4	67.2	71.1
un	un	84.1	51.72	64.0

Table 5: Cross-genre Name Tagging Performance

A and an organization B are represented as "*ARG1 (recipient)*" and "*ARG0 (agent)*" of a "*nominate-01*" event and B is identified as a "*political-party*", this relation can be aligned to the link by A's Wikipedia infobox "*political party*" (B). The Semantic Web community has spent a lot of manual efforts. Further improvement is likely to be obtained if such normalization can be automatically done using relation discovery and clustering techniques (Hasegawa et al., 2004) and techniques.

5.7.3 Linking Challenge: Morph

Another unique challenge in social media is resolving entity morphs which were created to avoid censorship, express strong sentiment or humor (Huang et al., 2013; Zhang et al., 2014). For example, "*Christie the Hutt*" is used to refer to "*Chris Christie*"; "*Dubya*" and "*Shrub*" are used to refer to "*George W. Bush*". Morph mentions usually share little surface features with their KB titles, and thus simple string matching methods fail to retrieve the correct candidates. On the other hand it's computationally intractable to search among all KB entries. Therefore, some

Year	Tasks & Resources	Methods	F-Measure	Example References
1966	-	First person name tagger with punch card 30+ decision tree type rules	-	(Borkowski et al., 1966)
1998	MUC-6	MaxEnt with diverse levels of linguistic features	97.12%	(Borthwick and Grishman, 1998)
2003	CONLL	System combination; Sequential labeling with Conditional Random Fields	89%	(Florian et al., 2003; McCallum et al., 2003; Finkel et al., 2005)
2006	ACE	Diverse levels of linguistic features, Re-ranking, joint inference	~89%	(Florian et al., 2006; Ji and Grishman, 2006)

Table 4: Name Tagging Milestones

global temporal distribution or social distance based methods that beyond text processing are required to address this problem.

5.7.4 Linking: Commonsense Knowledge

Some other remaining cases require common sense knowledge. For example, many approaches mistakenly linked "*William J. Burns*" in the following sentence '*During talks in Geneva attended by William J. Burns Iran refused to respond to Solana's offers*" to "*William J. Burns (1861-1932)*" in the KB. If we knew this source document was created in July 26, 2008 and so the entity was alive at that time since he was involved in the events such as attending talks, we would not have linked it to the dead person described in the KB. Such commonsense knowledge is very hard to acquire and represent. Recent work by Chen et al. (2014) attempted automatic discovery of commonsense knowledge for relation extraction. It may be also worth exploiting existing manually created commonsense knowledge from Concept Net (Liu and Singh, 2004) or FrameNet (Baker and Sato, 2003).

6 Cross-lingual Entity Linking

Only two teams HITS (Judea et al., 2014) and IBM (Sil and Florian, 2014) submitted runs to the Spanish-to-English cross-lingual entity linking task. Table 6 presents the B-cubed+ scores for these two systems. Both systems followed their English Entity Linking approaches. It's very encouraging to see that IBM system achieved similar performance with the top English EDL

Query	Team	B-cubed+ (%)		
		P	R	F
Spanish	HITS1	78.9	68.4	73.2
	IBM1	84.0	81.6	82.8
English	HITS1	68.4	60.3	64.1
	IBM1	80.6	77.7	79.1

Table 6: Cross-lingual Entity Linking Performance

system, although the difficulty level of queries are not comparable.

This year many teams expressed interests in participating the Chinese-to-English cross-lingual entity linking task, but they all ended up with focusing on English EDL task. In KBP2015 we will propose a new tri-lingual EDL task by extending the source collection from English only to three languages: English, Chinese and Spanish.

7 Conclusions and Future Work

The new EDL task has attracted many interests from the KBP community and produced some interesting research problems and new directions. In KBP2015 we will focus on the following extension and improvement:

- Improve the annotation guideline and annotation quality of the training and evaluation data sets;

- Develop more open sources, data and resources for Spanish and Chinese EDL;

- Encourage researchers to re-visit the entity mention extraction problem in the new cold-start KBP setting;

- Propose a new tri-lingual EDL task on a source collection from three languages: English, Chinese and Spanish;

- Investigate the impact of EDL on the end-to-end cold-start KBP framework.

Acknowledgments

References

I. Anastacio, B. Martins, and P. Calado. 2011. Supervised learning for linking named entities to knowledge base entries. In *Proc. Text Analysis Conference (TAC2011)*.

Amit Bagga and Breck Baldwin. 1998. Algorithms for scoring coreference chains. In *Proc. the LREC Workshop on Linguistic Coreference*, pages 563–566.

Collin F. Baker and Hiroaki Sato. 2003. The framenet data and software. In *Proc. ACL2003*.

L. Banarescu, C. Bonial, S. Cai, M. Georgescu, K. Griffitt, U. Hermjakob, K. Knight, P. Koehn, M. Palmer, and N. Schneider. 2013. Abstract meaning representation for sembanking. In *Proc. Linguistic Annotation Workshop*.

A. Barrena, E. Agirre, and A. Soroa. 2014. Ubc entity discovery and linking and diagnostic entity linking at tac-kbp 2014. In *Proc. Text Analysis Conference (TAC2014)*.

Razvan Bunescu. 2006. Using encyclopedic knowledge for named entity disambiguation. In *EACL*, pages 9–16.

Jie Cai and Michael Strube. 2010. Evaluation metrics for end-to-end coreference resolution systems. In *Proc. SIGDIAL*, pages 28–36.

T. Cassidy, H. Ji, L. Ratinov, A. Zubiaga, and H. Huang. 2012. Analysis and enhancement of wikification for microblogs with context expansion. In *Proceedings of COLING 2012*.

D. Ceccarelli, C. Lucchese, S. Orlando, R. Perego, and S. Trani. 2013. Learning relatedness measures for entity linking. In *Proceedings of the 22Nd ACM International Conference on Conference on Information & Knowledge Management*, CIKM '13.

Z. Chen and H. Ji. 2011. Collaborative ranking: A case study on entity linking. In *Proc. EMNLP2011*.

L. Chen, Y. Feng, S. Huang, Y. Qin, and D. Zhao. 2014. Encoding relation requirements for relation extraction via joint inference. In *Proc. the 52nd Annual Meeting of the Association for Computational Linguistics (ACL2014)*.

X. Cheng and D. Roth. 2013. Relational inference for wikification. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*.

Silviu Cucerzan. 2007. Large-scale named entity disambiguation based on wikipedia data. In *EMNLP-CoNLL 2007*.

S. Cucerzan. 2011. Tac entity linking by performing full-document entity extraction and disambiguation. In *Proc. TAC 2011 Workshop*.

M. Dojchinovski, I. Lasek, O. Zamazal, and T. Kliegr. 2014. Entityclassifier.eu and semitags: Entity discovery, linking and classification with wikipedia and dbpedia. In *Proc. Text Analysis Conference (TAC2014)*.

M. Dredze, P. McNamee, D. Rao, A. Gerber, and T. Finin. 2010. Entity disambiguation for knowledge base population. In *Proceedings of the 23rd International Conference on Computational Linguistics*.

J. Ellis, J. Getman, and S. M. Strassel. 2014. Overview of linguistic resource for the tac kbp 2014 evaluations: Planning, execution, and results. In *Proc. Text Analysis Conference (TAC2014)*.

N. Fernandez, J. A. Fisteus, L. Sanchez, and E. Martin. 2010. Webtlab: A cooccurence-based approach to kbp 2010 entity-linking task. In *Proc. TAC 2010 Workshop*.

P. Ferragina and U. Scaiella. 2010. Tagme: on-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedings of the 19th ACM international conference on Information and knowledge management*, CIKM '10.

J. Flanigan, S. Thomson, J. Carbonell, C. Dyer, and N. A. Smith. 2014. A discriminative graph-based parser for the abstract meaning representation. In *Proc. the 52nd Annual Meeting of the Association for Computational Linguistics (ACL2014)*.

T. Gao, J. R. Hullman, E. Adar, B. Hecht, and N. Diakopoulos. 2014. Newsviews: an automated pipeline for creating custom geovisualizations for news. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI2014)*.

Y. Guo, W. Che, T. Liu, and S. Li. 2011. A graph-based method for entity linking. In *Proc. IJCNLP2011*.

S. Guo, M. Chang, and E. Kiciman. 2013. To link or not to link? a study on end-to-end tweet entity linking. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Ben Hachey, Joel Nothman, and Will Radford. 2014. Cheap and easy entity evaluation. In *Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics (Vol. 2)*, pages 464–469.

X. Han and L. Sun. 2011. A generative entity-mention model for linking entities with knowledge base. In *Proc. ACL2011*.

X. Han and J. Zhao. 2009. Named entity disambiguation by leveraging wikipedia semantic knowledge. In *Proceedings of the 18th ACM conference on Information and knowledge management*, CIKM 2009.

X. Han, L. Sun, and J. Zhao. 2011. Collective entity linking in web text: A graph-based method. In *Proc. SIGIR2011*.

T. Hasegawa, S. Satoshi Sekine, and R. Grishman. 2004. Discovering relations among named entities from large corpora. In *Proc. the Annual Meeting of Association of Computational Linguistics (ACL 04)*.

J. Hoffart, M. Yosef, I. Bordino, H. Fürstenau, M. Pinkal, M. Spaniol, B. Taneva, S. Thater, and G. Weikum. 2011. Robust disambiguation of named entities in text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11.

Y. Hong, X. Wang, Y. Chen, J. Wang, T. Zhang, J. Zheng, D. Yu, Q. Li, B. Zhang, H. Wang, X. Pan, and H. Ji. 2014. Rpi-soochow kbp2014 system description. In *Proc. Text Analysis Conference (TAC2014)*.

H. Huang, Z. Wen, D. Yu, H. Ji, Y. Sun, J. Han, and H. Li. 2013. Resolving entity morphs in censored data. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

H. Huang, Y. Cao, X. Huang, H. Ji, and C. Lin. 2014a. Collective tweet wikification based on semi-supervised graph regularization. In *Proc. the 52nd Annual Meeting of the Association for Computational Linguistics (ACL2014)*.

H. Huang, Y. Cao, X. Huang, H. Ji, and C.Y. Lin. 2014b. Collective tweet wikification based on semi-supervised graph regularization. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).

H. Ji and D. Lin. 2009. Gender and animacy knowledge discovery from web-scale n-grams for unsupervised person mention detection. In *Proc. PACLIC 2009*.

H. Ji, R. Grishman, and H. T. Dang. 2011. An overview of the tac2011 knowledge base population track. In *Proc. Text Analysis Conf. (TAC'11)*, Gaithersburg, Maryland, Nov.

A. Judea, B. Heinzerling, A. Fahrni, and M. Strube. 2014. Hits' monolingual and cross-lingual entity linking system at tac 2014. In *Proc. Text Analysis Conference (TAC2014)*.

Z. Kozareva, K. Voevodski, and S. Teng. 2011. Class label enhancement via related instances. In *Proc. EMNLP2011*.

S. Kulkarni, A. Singh, G. Ramakrishnan, and S. Chakrabarti. 2009. Collective annotation of wikipedia entities in web text. In *KDD*.

Q. Li, H. Li, H. Ji, W. Wang, J. Zheng, and F. Huang. 2012. Joint bilingual name tagging for parallel corpora. In *Proc. 21st ACM International Conference on Information and Knowledge Management (CIKM'12)*, Maui, Hawaii, Oct.

H. Liu and P. Singh. 2004. Conceptnet - a pratical commonsense reasoning toolkit. 22:211–226.

X. Liu, Y. Li, H. Wu, M. Zhou, F. Wei, and Y. Lu. 2013. Entity linking for tweets. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Xiaoqiang Luo. 2005. On coreference resolution performance metrics. In *Proc. HLT/EMNLP*, pages 25–32.

K. Mazaitis, R. C. Wang, B. Dalvi, and W. W. Cohen. 2014. A tale of two entity linking and discovery systems. In *Prof. Text Analysis Conference (TAC2014)*.

P. McNamee and H.T. Dang. 2009. Overview of the tac 2009 knowledge base population track. In *Text Analysis Conf. (TAC) 2009*.

P. McNamee, J. Mayfield, D. Lawrie, D. W. Oard, and D. Doermann. 2011. Cross-language entity linking. In *Proc. IJCNLP2011*.

E. Meij, W. Weerkamp, and M. de Rijke. 2012. Adding semantics to microblog posts. In *Proceedings of the fifth ACM international conference on Web search and data mining*, WSDM '12.

M. Meurs, H. Almeida, L. Jean-Louis, and E. Charton. 2014. Semlinker entity discovery and linking system at tac kbp2014. In *Proc. Text Analysis Conference (TAC2014)*.

R. Mihalcea and A. Csomai. 2007. Wikify!: linking documents to encyclopedic knowledge. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, CIKM '07.

D. Milne and I.H. Witten. 2008. Learning to link with wikipedia. In *Proceeding of the 17th ACM conference on Information and knowledge management*, pages 509–518. ACM.

S. Monahan, D. Carpenter, M. Gorelkin, K. Crosby, and M. Brunson. 2014. Populating a knowledge base with entities and events. In *Proc. Text Analysis Conference (TAC2014)*.

T. H. Nguyen, Y. He, M. Pershina, X. Li, and R. Grishman. 2014. New york university 2014 knowledge base population systems. In *Proc. Text Analysis Conference (TAC2014)*.

Joel Nothman, Nicky Ringland, Will Radford, Tara Murphy, and James R. Curran. 2013. Learning multilingual named entity recognition from Wikipedia. *Artificial Intelligence*, 194:151–175, January.

M. Pennacchiotti and P. Pantel. 2009. Entity extraction via ensemble semantics. In *Proc. EMNLP2009*.

W. Radford, B. Hachey, J. Nothman, M. Honnibal, and J. R. Curran. 2010. Cmcrc at tac10: Document-level entity linking with graph-based re-ranking. In *Proc. TAC 2010 Workshop*.

Lev Ratinov and Dan Roth. 2009. Design challenges and misconceptions in named entity recognition. In *Proc. CONLL*, pages 147–155.

L. Ratinov and D. Roth. 2012. Learning-based multi-sieve co-reference resolution with knowledge. In *EMNLP*.

L. Ratinov, D. Roth, D. Downey, and M. Anderson. 2011. Local and global algorithms for disambiguation to wikipedia. In *Proc. of the Annual Meeting of the Association of Computational Linguistics (ACL)*.

W. Shen, J. Wang, P. Luo, and M. Wang. 2013. Linking named entities in tweets with knowledge base via user interest modeling. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13.

A. Sil and R. Florian. 2014. The ibm systems for english entity discovery and linking and spanish entity linking at tac 2014. In *Proc. Text Analysis Conference (TAC2014)*.

A. Sil and A. Yates. 2013. Re-ranking for joint named-entity recognition and linking. In *Proc. CIKM2013*.

Y. Song and D. Roth. 2014. On dataless hierarchical text classification. In *Proc. AAAI 2014*.

S. Tamang, Z. Chen, and H. Ji. 2012. Cuny_blender tac-kbp2012 entity linking system and slot filling validation system. In *Proc. Text Analytics Conf. (TAC'12))*, Gaithersburg, Maryland, Nov.

D. Vitale, P. Ferragina, and U. Scaiella. 2012. Classification of short texts by deploying topical annotations. In *ECIR*, pages 376–387.

B. Zhang, H. Huang, X. Pan, H. Ji, K. Knight, Z. Wen, Y. Sun, J. Han, and B. Yener. 2014. Be appropriate and funny: Automatic entity morph encoding. In *Proc. the 52nd Annual Meeting of the Association for Computational Linguistics (ACL2014)*.

G. Zhao, P. Lv, R. Xu, Q. Zhi, and J. Wu. 2014. Msiipl thu at tac 2014 entity discovery and linking track. In *Proc. Text Analysis Conference (TAC2014)*.

Z. Zheng, F. Li, M. Huang, and X. Zhu. 2010. Learning to link entities with knowledge base. In *Proceedings of NAACL-HLT2010*.

J. G. Zheng, D. Howsmon, B. Zhang, J. Hahn, D. McGuinness, J. Hendler, and H. Ji. 2014. Entity linking for biomedical literature. In *Proc. CIKM2014 Workshop on Data and Text Mining in Biomedical Informatics*.

S. Zhou, C. Kruengkrai, and K. Inui. 2014. Tohoku's entity discovery and linking system at tac 2014. In *Prof. Text Analysis Conference (TAC2014)*.