Challenges in the Knowledge Base Population Slot Filling Task

Bonan Min and Ralph Grishman
New York University
715 Broadway, 7th floor
New York, NY 10003, USA
{min, grishman}@cs.nyu.edu

Abstract

The Knowledge Based Population (KBP) evaluation track of the Text Analysis Conferences (TAC) has been held for the past 3 years. One of the two tasks of KBP is slot filling: finding within a large corpus the values of a set of attributes of given people and organizations. This task has proven very challenging, with top systems rarely exceeding 30% F-measure. In this paper, we present an error analysis and classification for those answers which could be found by a manual corpus search but were not found by any of the systems participating in the 2010 evaluation. The most common sources of failure were limitations on inference, errors in coreference (particularly with nominal anaphors), and errors in named entity recognition. We relate the types of errors to the characteristics of the task and show the wide diversity of problems that must be addressed to improve overall performance.

Keywords: Information Extraction, Evaluation, Analysis

1. Introduction

The Knowledge Based Population (KBP)[1] track of the Text Analysis Conferences (TAC) organized by the U.S. National Institute of Standards and Technology (NIST) has been held every year since 2009. It attracts a large number of participants[2]. As the successor to the Message Understanding Conference (MUC)[3] and Automatic Content Extraction (ACE)[4] evaluations, KBP removes some artificial constraints in the task definition and uses a large set of documents which have not gone through a careful selection process. This has made the task more realistic and very challenging. For the Slot Filling task which we will discuss in this paper, top systems rarely exceed 30% F-measure.

To understand what the challenges of the Slot Filling task are, we generated a list of answers which are found by human annotators but none of the systems in the 2010 evaluation, and then manually checked the sources of error and categorized them. We hope this analysis can help researchers to understand the problem better and thus enable us to find better solutions.

2. The TAC-KBP Slot Filling task

2.1 Task definition

The KBP track at TAC 2010 provides a corpus of 1.7 million newspaper articles and weblog/newsgroup posts and an initial knowledge base (KB) based on Wikipedia infoboxes. Two KBP tasks are defined: Entity Linking (given a name and a document in the corpus containing that name, to identify the KB entry corresponding to that name, if it exists) and Slot Filling. In the Slot Filling task, participant systems are given a list of person and organization names (the 'queries'), must locate information about these entities in the corpus and then update the KB by filling empty attributes ('slots') with correct, non-redundant information and at the same time, provide the reference articles that justify the filled information. There are 42 slots – 26 for person entities and 16 for organization entities. Example slots include *per:employee_of* and *org:top_members/employees*. Some slots (such as *per:date_of_birth*) can only take on a single value, while others (such as *per:member_of*) can take multiple values. A small set of queries and their slot values are provided as training data. More information can be found in the task definition document of KBP 2010[5].

2.2 Evaluation

Given a list of 100 queries (names along with example articles for disambiguating the names), each site is expected to find answers for the slots of the queries. Evaluation for precision is relatively straightforward -- LDC (Linguistic Data Consortium) annotators assess whether the answers are correct by manually checking with the reference article. Evaluation of true recall is almost impossible -- the corpus is too big for a human to read through to find all the answers. Therefore KBP adopts a pseudo-recall evaluation, in which the LDC annotators pool all system outputs, assess them, and use the correct answers in the pool as the complete list for recall calculation. To make the answer list more complete, LDC conducts a separate human slot filling round, in which LDC annotators use internal tools to find related articles, and then use their own judgment to find answers within these articles. These answers are also put into the answer pool for adjudication.

[1] http://nlp.cs.qc.cuny.edu/kbp/2010/

[2] 23 participants submitted their results to one or both of the Entity Linking and Slot Filling tasks in KBP 2010.

[3] http://www-nlpir.nist.gov/related_projects/muc

[4] http://www.itl.nist.gov/iad/mig/tests/ace/

[5] http://nlp.cs.qc.cuny.edu/kbp/2010/KBP2010_TaskDefinition.pdf

2.3 Overview of Participating Systems

The KBP participants have applied a wide range of techniques to tackle the Slot Filling problem (Ji et al. 2010; Ji and Grishman 2011). The basic flow is similar: most systems use a **passage retrieval** module to retrieve related documents or passages, and then run an **answer extraction** module to find answers, followed by methods to validate and merge responses. Figure 9 in (Ji et al. 2010) shows the common Slot Filling system architecture. Here we briefly summarize the key techniques used in participating systems, categorized by modules or the targeted problems:

Passage retrieval: Since the million-level document corpus provided by KBP is too large to be processed at runtime, most systems use an IR engine to retrieve related documents or passages that may contain answers for further processing. To provide better search terms, participants, e.g., IBM (Castelli et al., 2010) and CUNY (Chen et al., 2010), usually perform query expansion, such as acronym restoration with the example article that comes with the query, or expand queries with alternative names extracted from Wikipedia redirect links. Participants such as IBM also indexed entity types and KB entries to support search by specific entity types or KB node IDs.

Answer extraction: some systems, e.g., IIRG (Byrne and Dunnion, 2010) and CUNY, adapt Question Answering (QA) techniques. With QA, values in the retrieved documents/passages that have the correct targeted types and highest confidence score or frequency (IIRG) are returned (for multi-value slots, several high confidence answers are kept). Other systems use information extraction (IE) techniques. There are two popular approaches: some participants, e.g. IBM, CUNY and LCC (Lehmann et al. 2010) use supervised algorithms for relation extraction. This normally involves entity mention detection, coreference resolution and then relation detection and classification. Participants use hand-written rules to map their system outputs to each slot type. Some participants (NYU (Grishman and Min 2010) and CUNY) apply a large set of patterns, obtained with bootstrapping, for answer extraction. Learning extraction patterns from a bootstrapping procedure has the advantage of not requiring lots of training data (only a small number of seeds), but needs to be cautious to prevent semantic drift.

Source of training: KBP provides a limited amount of training data; for the 2010 evaluation, around 150 queries with filled attribute values were provided. In order to obtain more labeled data for training a supervised relation extractor, a few approaches were tried: IBM built annotated data under its IE annotation framework KLUE (Han, 2010), CUNY used the relation extractor in JET (Grishman et al., 2005) trained from ACE training data, while LCC used active learning to rapidly construct training data for KBP slots. In contrast to using the relatively high quality training data, the Stanford team (Surdeanu et al., 2010) heuristically aligned relational tables to unstructured text to automatically construct training data for the purpose. Recent work (Riedel et al., 2010) shows that this distant supervision approach needs a sophisticated learning algorithm to tolerate noise. This method makes use of external knowledge bases, such as Wikipedia infoboxes and Freebase.

Features for IE systems: For supervised IE modules which extract answers from local clues, mainly sentence-level features are used. For example, the IBM system used lexical, syntactic and semantic features.

Coreference resolution: most systems use cross-sentence coreference resolution in conjunction with the answer extraction module. Some sites, e.g. IBM (Castelli et al. 2010), which participated or previously participated in the Entity Linking task of KBP, did cross-document coreference for entity resolution, which helps redundancy removal among the answers.

Inference: as we will show later, KBP requires systems to perform quite sophisticated inference. IBM and CUNY apply predefined inference rules between slots (relations and events). In particular, IBM's rules include rules operating on coreferential entity mentions, rules between KLUE relations and rules between slots and KLUE relations. Some high-confidence long-distance inference cases can be handled by recursively applying these rules.

Implicit arguments: NYU handles implicit arguments for corporate titles not explicitly tied to an organization by assuming that the organization (if it exists and is unique) from the previous sentence is the implicit argument of the title.

System combination: a few participants take advantage of combining different extraction techniques. For example, CUNY combines systems based on IE and QA techniques for answer extraction.

3. Challenges in the KBP Slot Filling task

Taken together, these systems incorporate most of the methods currently being used for IE. Slot fills which were missed by all the systems are therefore likely, for the most part, to reflect the limitations of current IE technology. To understand these limitations, we gathered the answers that human annotators could find but none of the systems got, and then identified their likely sources of error. These sources are the major challenges for the KBP Slot Filling task.

Based on the system output of all participants, the LDC answers found by manual search and the assessment file which contains the manual assessment of all pooled answers (included both system outputs and LDC answers), we took the following preprocessing steps to generate the answer list of interest:

1. Generate a list of equivalence classes of answers (for scoring purposes, annotators group responses which refer to the same entity, such as "Bill" and "William", into equivalence classes).
2. For each correct answer in the list of LDC answers, if there is no answer from the same equivalence class for the slot in the merged list of participant answers, put it into the final list.

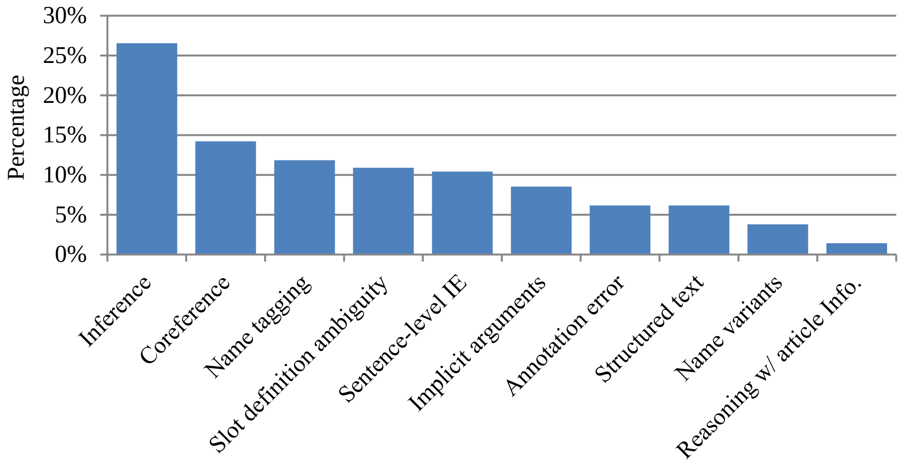

Figure 1: Sources of errors

After these two steps, we obtain a list of correct answers which none of the participants get but the human annotators can successfully identify. It contains 140 answers.

We manually analyzed all answers by reading the reference article, looking for evidence from which the annotators drew their conclusions. Based on our experience in Information Extraction and the system descriptions from participants, we categorized these answers by their likely sources of error, as shown in Figure 1, in some cases listing multiple sources. We further categorized each source into sub-categories and illustrate each one with an example, as shown in Table 1.

Figure 1 shows that requiring *inference* is the most common source of error, followed by difficult *coreference* cases, hard *name tagging* cases, and so on.

Inferences Some examples require quite sophisticated reasoning involving facts scattered in the documents and in world knowledge. For the example of *org:members (National Christmas Tree Association, River Ridge Tree Farms)* in Table 1, we need the following inferential rules for extracting the relation:

host_of(National Christmas Tree Association(NCTA), National Christmas Tree Contest(NCTC)) ^
eligible_for_attending(members of NCTA, NCTC) ^
attended(winner, NCTC) ^ win(River Ridge, NCTC) ^
coreferential(River Ridge Tree Farms, River Ridge) →
org:members(NCTA, River Ridge Tree Farms)

To find this answer, a system must be able to extract several domain-dependent relations and events (*host_of*, *eligible_for attending*, and *win*) from the text, and integrate common knowledge (the winner must have attended the contest) for complicated reasoning. Systems are also required to understand event causality for some slots (as shown in Table 1).

Coreference Coreference failure was the second most frequent source of error. Problems in resolving nominal anaphors accounted for around two-thirds of the cases (19 out of 30), followed by difficult pronoun anaphors, and some cases with collective nouns as anaphors. An example of nominal anaphors is shown in the following text:

the alleged prostitution outfit, known as Pamela Martin and Associates, that she is accused of running by phone out of her homes in Vallejo and Escondido, Calif. ...The operation, …

The system is expected to understand that “the alleged prostitution outfit”, “Pamela Martin and Associates”, and “The operation” are coreferential.

Furthermore, it is even harder if a system needs to perform resolution and understand document structure at the same time. The example in Table 1, *per:schools_attended(Erika Rose, Boston College)* is a case of pronoun coreference in semi-structured text.

Name tagging is a critical component for traditional IE systems, and it continues to be critical for the Slot Filling task. Difficult examples include *I'm a Celebrity, Get Me Out of Here* (television show name) and *The Rising* (band name). This requires the name tagger to handle the wide range of domains that are covered by the KBP corpus.

Sentence-level IE refers to the sentence-level relation extraction problem studied in previous IE evaluations such as ACE. Given a sentence that contains both the query and the answer, a system is expected to detect and characterize the relation between them, if the sentence contains expressions indicating the relation. Inability to identify such relations accounts for around 10% of the failures. In most cases, the relation is expressed through a complicated long-range linguistic structure within a sentence.

Implicit arguments In some cases slots can be filled only by recognizing implicit relations between predicates and their arguments (“bridging”). Failures to capture such arguments account for around 8.5% of the examples. This shows its importance for the Slot Filling task. For the example in Table 1, the system is expected to understand that *vice chairman* and *senior group executive* are the top employees with *Samsung*, which appeared previously in the document but was omitted in the sentence containing the answer.

Structured text Slot Filling also requires systems to deal with structured text (where relations are expressed through page layout or markup). For example (as shown in Table 1), some answers appeared in dialog transcripts, in which a system should figure out the speaker for each

Sources of error		Examples	
Categories	Sub-categories	Slots, queries and answers	Sampled reference text (unrelated sentences are omitted and replaced with "…" to save space)
Inference	cross-slot (relations or events)	org:members (*National Christmas Tree Association, River Ridge Tree Farms*)[6]	*Jessie Davis and Russell Estes, owners of* ***River Ridge Tree Farms*** *in Crumpler, North Carolina, where the tree was grown, joined the first lady, along with their families….The* ***National Christmas Tree Association*** *has presented the official White House tree since 1966..... Members of the association compete in state and regional competitions to become eligible to take a tree to the national contest. River Ridge was named grand champion in the National Christmas Tree Contest in August.*
	event causality	per:cause_of_death (*Michael Sandy, chased into the path of a moving car*)	*Jurors deliberated several days before convicting Anthony Fortunato in the death of* ***Michael Sandy****, a gay man who was beaten and then* ***chased into the path of a moving car*** *on Brooklyn's Belt Parkway on Oct. 8, 2006.*
	long distance reasoning	org:parents(*Nitschmann Middle School, Bethlehem Area School District*)	*John Acerra, 50, of Allentown, Pennsylvania, was arrested Tuesday in his office at* ***Nitschmann Middle School*** *in Bethlehem, where police said they found meth on his desk....A letter was sent to parents informing them of Acerra's arrest and teachers had spoken with the school's 950 students,* ***Bethlehem Area School District*** *Superintendent Joseph Lewis told reporters Wednesday.*
	with background knowledge	org:subsidiaries (*Massachusetts House of Representatives, Joint Committee on the Environment, Natural Resources, and Agriculture*)	*Representative Frank Smizik, a Brookline Democrat who is House Large wind farms could be constructed in state waters under legislation passed by the* ***Massachusetts House of Representatives*** *Wednesday that critics said could aid a controversial wind energy project in Buzzards Bay....chairman of the* ***Joint Committee on the Environment, Natural Resources, and Agriculture****, said...*
Coreference	nominal	org:city_of_headquarters (*Institute for Diversity and Ethics in Sport, Orlando*)	*The NFL is the only U.S. pro sports organization that refuses to share its league office data with University of Central Florida's* ***Institute for Diversity and Ethics in Sports****, which also conducts annual studies on the NBA, Major League Baseball....Richard Lapchick, report author and head of UCF's diversity institute in* ***Orlando****, Fla., said the league data would probably be better than the NFL's most recent grade.*
	pronoun	per:stateorprovinces_of_residence(*Ezra Levant, Alberta*)	*Alberta Human Rights Commission Interrogation Opening remarks by* ***Ezra Levant****, January 11, 2008 – Calgary My name is Ezra Levant....But I do have faith in the justice and good sense of my fellow* ***Albertans*** *and Canadians.*
	collective nouns	per:employee_of(*Holly Montag, NBC*)	*Life & Style is reporting that the Holly/Sanjaya romance that* ***Holly Montag*** *recently denied is totally on. The two recently competed/starred on* ***NBC****'s* ***I'm a Celebrity, Get Me Out of Here****.*
Name tagging			
Slot definition ambiguity		org:top_members/employees (*Massachusetts House of Representatives, Edward J. Markey*)	*WASHINGTON -* ***Massachusetts****' 10* ***House*** *members have emerged as key tacticians and advisers to House Speaker Nancy Pelosi…Representative* ***Edward J. Markey of Malden****, …*
Sentence-level IE		per:cities_of_residence(*Michael Johns, Buckhead*)	***Michael Johns*** *(29) Currently lives in Los Angeles, but was born in Perth, Australia. Johns moved to the U.S. in 1998 to attend Abraham Baldwin Agriculture College in Tifton, GA, then moved to* ***Buckhead****, GA outside of Atlanta to pursue singing.*
Implicit arguments		org:top_members/employees(*Samsung, Kim In-Joo*)	*Special prosecutors said Thursday they have charged* ***Samsung*** *chairman Lee Kun-Hee …. pending trial as would nine other executives who were also charged. These include vice chairman Lee Hak-Soo and senior group executive* ***Kim In-Joo****.*
Annotation error	wrong answer	org:website(*Pamela Martin and Associates , http://www.deborahjeanepalfrey.com*)	*In court records, prosecutors estimate that her business,* ***Pamela Martin and Associates****, generated more than $2 million (euro1.5 million) in revenue over 13 years,….* *On the Net:* ***http://www.deborahjeanepalfrey.com***
	over reasoning	org:founded(*Ownit Mortgage Solutions, 2003*)	*Visitors to* ***Ownit Mortgage Solutions****' offices here are met by an unmanned reception desk....allas created Ownit from a small mortgage company he and his partners bought in* ***2003*** *for $30 million.*
Structured text	dialog	per:schools_attended (*Erika Rose, Boston College*)	*[interviewer] …* *[interviewee, Erika Rose] I was studying Economics at* ***BostonCollege***

[6] In which org:members is the slot name, *National Christmas Tree Association* is the query and *River Ridge Tree Farms* is the slot value.

	tables, list or forms	per:date_of_birth(*Lewis Hamilton, 1985-01-07*)	*Biography of the 2008 F1 champion **Lewis Hamilton**...* *Lewis Hamilton* *BORN: **Jan. 7, 1985**, Stevenage, England.*
Name variants (including a few cases in which the strings are not exactly names)	answer normalization	per:age(*Molly Malaney, 24*)	*...to Jason's feelings for **Molly Malaney** prior to his tearful TV confession...Jason needs some help. He picked a **24** yr. old girl (Molly) …*
	phonetic name variants	per:alternate_names (*Kate Gosselin, K8*)	*Well folks, **Kate Gosselin**'s gone and fancied up her lady mullet again.…I'm sorry, **K8** (GET IT? BECAUSE SHE HAS EIGHT KIDS! I'M SUCH A SCAMP), but since when is plopping on Britney's weave from a year ago considered starting over?*
	Query name variants	org:alternate_names(*Norris Comprehensive Cancer Center, Norris Cancer Center*)	*The debate over Provenge highlights how difficult it is for scientists and the FDA to reach a decision about some drugs, says Dr. David Penson, an associate professor of urology and preventive medicine at the Keck School of Medicine at the University of Southern California/**Norris Cancer Center**.*
Reasoning with article information		per:date_of_birth(*Chante Moore, 1967-02-17*)	*[Note: The document is created in 2007. It is list of celebrities whose birthday is in Febrary]* *Feb. 17: ….Singer **Chante Moore** is **40**.*

Table 1. Detailed sources of error with examples. Text in [] does not appear in the original document, but is included to clarify the examples.

turn and resolve first-person pronouns to the correct entity, then allow the answer extraction module to extract answers with linguistic clues. Some text (especially from weblogs) contains tabular information that has lost its physical form. Such cases are hard for systems to extract answers from. Sometimes the tabular information causes errors. CUNY applied a few case-dependent rules to filter incorrect answers extracted from such text.

Name variants Name variants, mainly query and answer variants, account for 3.8% errors. Most systems deal with this problem by looking up alternative names from a large dictionary of names and their variants, or applying more sophisticated Entity Linking techniques to link variants to base forms. Some examples are particularly hard. For example, the phonetic name variant *K8* is a name variant for *Kate*, shown in Table 1.

R**easoning with article information** (metadata): in order to fill some slots, the systems are also required to do reasoning with the article information. For example, a document contains *Feb. 17: ….Singer Chante Moore is 40.* The system should conclude that the birthdate of *Chante Moore* is 40 years ahead of the document creation date.

There are two sources of error that are related to the evaluation itself. The first one is **annotation error**, by which we mean that annotators made a mistake according to the annotation guidelines. We found that for around 6% of the examples, the annotators either find a *wrong answer* (but the answer is marked correct during the adjudication), or make unwarranted inferences (*over reasoning*) when looking for the answer. For example, in *org:founded(Ownit Mortgage Solutions, 2003)*, the founders bought a small company in 2003. This doesn't necessarily mean that they founded the new company in the same year. The second source of error is **slot definition ambiguity**, by which we mean that the annotation guidelines don't specify whether this filler should be considered a correct answer. For example, whether representatives should be considered *top_members/employees* of the state *House of Representatives* is not very clear. This source of error accounts for 10% of the examples. These ambiguities should be resolved in the annotation guidelines for further KBP evaluations; a few were resolved for 2011.

3.1 Discussion

Surprisingly, the most studied problem, which is lack of alternate linguistic expressions of each attribute, is not the dominant problem for the KBP Slot Filling task. Listed as **sentence-level IE** in figure 1, this source of error only accounts for around 10% of the failures.

The analysis in the previous section shows that the KBP Slot Filling task expects a system to address all the problems listed in addition to providing a good core extraction system. The system should also be able to handle a heterogeneous corpus, for example, with both unstructured and semi-structured text.

Furthermore, we observe that around 39% of examples have 2 sources of error and around 13% have 3 sources of error. For example, a system is expected to correctly resolve pronouns in a dialog whose structure is encoded by layout. This shows that, in order to achieve good performance, a system should be able to solve these problems jointly rather than deal with each of them independently.

4. Relations to other evaluations

There are several evaluations that are related to KBP Slot Filling. However, the KBP Slot Filling task raises different challenges.

Compared to previous IE evaluations such as MUC and ACE, KBP provides a much larger and less curated corpus, which covers a wide range of topics. There are 4 major differences:

- The KBP 2010 corpus contains 1.7 million articles, which is significantly larger than other IE evaluations. It requires a system to find answers from the entire corpus rather than individual documents. This is more realistic and challenging than previous IE evaluations such as MUC and ACE.

- The KBP corpus contains plain-text articles of different formats, e.g, news articles, weblogs and newsgroup posts. It also contains some less clean articles, thus raising additional problems for existing systems. For previous evaluations such as MUC and ACE, test documents were hand-selected, generally avoiding documents where page layout provided crucial information. By using a less curated corpus, the KBP Slot Filling evaluation exposes the brittleness of some strategies.
- Articles in the KBP corpus are from a wide range of topics, thus a good system should not be crafted for a specific domain.
- Slot Filling does not require the answer to appear in the same sentence with the query; this makes the task harder than ACE relation extraction. ACE required the two arguments of a relation to be explicitly mentioned in the same sentence, reducing the need to recover implicit arguments or make inferences from document metadata.

A number of recent evaluations, such as DARPA Machine Reading [7] (Strassel et al. 2010), Question Answering for Machine Reading Evaluation (QA4MRE)[8] and Recognizing Textual Entailment (RTE)[9], have placed an explicit emphasis on inference. This has not been an explicit focus of Slot Filling. However, as we have shown in the previous section, inference plays an important role in achieving good KBP Slot Filling performance. Compared to the Machine Reading evaluations, the inference problem seems to be harder because of the breadth of the corpus; progress in Slot Filling may require the creation and integration of reasoning specialists for a number of common inference tasks.

In summary, the KBP Slot Filling task is designed as a more realistic task with fewer artificial constraints than previous tasks. Therefore the distribution of sources of errors is more meaningful for building a working system for extracting facts from text corpora.

5. Conclusion

In this paper, we analyzed examples from KBP assessment data that are challenging for current Slot Filling systems. This analysis does not point to a single dominant problem, but rather a wide range of different challenges which all need to be addressed to create a successful extraction system for more realistic data.

6. References

Lorna Byrne and John Dunnion. 2010. UCD IIRG at TAC 2010. *Proc. Text Analysis Conference (TAC 2010).*

Vittorio Castelli, Radu Florian and Ding-jung Han. 2010. Slot Filling through Statistical Processing and Inference Rules. *Proc.Text Analysis Conference (TAC 2010).*

Zheng Chen, Suzanne Tamang, Adam Lee, Xiang Li, Wen-Pin Lin, Matthew Snover, Javier Artiles, Marissa Passantino and Heng Ji. 2010. CUNYBLENDER TAC-KBP2010 Entity Linking and Slot Filling System Description. *Proc. Text Analysis Conference (TAC 2010).*

Ralph Grishman, David Westbrook and Adam Meyers. 2005. NYU's English ACE 2005 System Description. *Proc. Automatic Content Extraction (ACE) 2005 Workshop.*

Ralph Grishman and Bonan Min. 2010. New York University KBP 2010 Slot-Filling System. *Proc. Text Analysis Conference (TAC 2010).*

Ding-Jung Han. 2010. KLUE annotation guidelines - version. 2.0. Technical Report RC25042, IBM Research.

Heng Ji and Ralph Grishman. 2011. Knowledge Base Population: Successful Approaches and Challenges. *Proc. 49th Annual Meeting Assn. Computational Linguistics (ACL 2011).*

Heng Ji, Ralph Grishman, and Hoa Trang Dang. 2011. Overview of the TAC2011 Knowledge Base Population Track. *Proc. Text Analysis Conference (TAC 2011).*

Heng Ji, Ralph Grishman, Hoa Trang Dang, Kira Griffitt and Joe Ellis. 2010. An Overview of the TAC 2010 Knowledge Base Population Track. *Proc. Text Analysis Conference (TAC2010).*

John Lehmann, Sean Monahan, Luke Nezda, Arnold Jung and Ying Shi. 2010. LCC Approaches to Knowledge Base Population at TAC 2010. *Proc. Text Analysis Conference (TAC 2010).*

Sebastian Riedel, Limin Yao and Andrew McCallum. 2010. Modeling Relations and Their Mentions without Labeled Text. *Proc. European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD '10).*

Stephanie Strassel, Dan Adams, Henry Goldberg, Jonathan Herr, Ron Keesing, Daniel Oblinger, Heather Simpson, Robert Schrag, and Jonathan Wright. 2010. The DARPA Machine Reading program - encouraging linguistic and reasoning research with a series of reading tasks. In *Proceedings of LREC 2010*.

Mihai Surdeanu, Sonal Gupta, John Bauer, David McClosky, Angel X. Chang, Valentin I. Spitkovsky, Christopher D. Manning. 2010. Stanford's Distantly-Supervised Slot-Filling System. *Proc. Text Analysis Conference (TAC 2010).*

[7] http://www.darpa.mil/Our_Work/I2O/Programs/Machine_Reading.aspx

[8] http://celct.fbk.eu/QA4MRE/

[9] http://www.nist.gov/tac/2011/RTE/