Linguistic Resources for 2013 Knowledge Base Population Evaluations

Joe Ellis, Jeremy Getman, Justin Mott, Xuansong Li, Kira Griffitt, Stephanie M. Strassel, & Jonathan Wright

Linguistic Data Consortium
University of Pennsylvania
Philadelphia, PA 19104
U.S.A

Email: {joellis, jgetman, jmott, xuansong, kiragrif, strassel, jdwright} @ldc.upenn.edu

Abstract

Knowledge Base Population (KBP) is an evaluation track of the Text Analysis Conference (TAC), a workshop series organized by the National Institute of Standards and Technology (NIST). In 2013, the KBP evaluations included five tasks targeting information extraction and question answering technologies: Entity Linking, Slot Filling, Temporal Slot Filling, Sentiment Slot Filling, and Cold Start. The Sentiment and Temporal Slot Filling tasks were introduced in 2013 in an effort to move the KBP challenges into new domains, specifically beliefs and events. Linguistic Data Consortium (LDC) has supported the TAC KBP evaluation since 2009, each year producing new linguistic resources including data, annotations, system assessments, tools and specifications. This paper describes the resource creation efforts in support of TAC KBP 2013, with an emphasis on procedures and methodologies for query selection, annotation, and assessment.

1. Introduction

The Text Analysis Conference (TAC) is a series of evaluation workshops initiated by the National Institute of Standards and Technology (NIST) that aim to advance natural language processing technologies and applications. Knowledge Base Population (KBP), one of the on-going TAC tracks, started in 2009 with a focus on information extraction and question answering technologies. Evolved from the TREC Question Answering (Dang et al. 2006) and Automated Content Extraction (ACE) (Doddington et al. 2004) evaluation programs (McNamee et al. 2010), TAC KBP evaluates computation systems on five main tasks: Entity Linking, Slot Filling, Temporal Slot Filling, Sentiment Slot Filling, and Cold Start.

The Entity Linking task requires systems to either accurately link named mentions of person (PER), organization (ORG), and geopolitical (GPE) entities in text to entries in an external knowledge base, or correctly report if there are no matching entries. Entity Linking evaluations started in 2009 with an English only version (Simpson et al., 2010) and added cross-lingual Chinese and Spanish versions of the task in 2011 and 2012 respectively, both of which were continued in 2013. The Slot Filling task requires systems to automatically populate Wikipedia-style infoboxes for a set of specific named person (PER), and organization (ORG) entities with information retrieved from a collection of natural language English source documents. In 2012, Spanish Slot Filling queries and annotations were developed in an effort to move the task into cross-lingual terrain. However, a Spanish Slot Filling evaluation has not yet been conducted. Cold Start requires systems to construct a new knowledge base from the information contained in an unstructured text collection, effectively coordinating the separate technologies developed for Entity Linking and Slot Filling. In Sentiment Slot Filling, one of the new TAC KBP evaluations conducted for 2013, systems and annotators extract positive or

negative sentiments stated by or about query entities. Lastly, for the Temporal Slot Filling task, which was initially piloted in 2011 but not revived until 2013, performers seek to add temporal constraints on specified Slot Filling relations.

Linguistic Data Consortium (LDC) at the University of Pennsylvania has supported KBP evaluations since 2009 by creating and distributing linguistic resources including data, annotations, system assessment, tools and specifications. This paper describes the resource creation effort for 2013 TAC KBP. Section 2 describes the source data and knowledge base used for all KBP tasks; section 3 discusses the training and evaluation data provided by LDC for the 2013 KBP tasks; section 4 discusses procedures and methodologies for query selection, annotation, and assessment; and section 5 concludes the paper.

2. Source Data & Reference Knowledge Base

2013 saw the retirement of the original seed corpus used in TAC KBP from 2009 – 2012, TAC 2010 KBP Source Data (LDC2010E12). Obviously, the collection is still useful for the purpose of utilizing existing training data and so it was made available to participants in 2013; however, it was not used in any new data creation efforts.

Also for 2013, LDC developed a single package that included all of the source data for the evaluations (with the exception of Cold Start, for which the corpus had to remain hidden until the time of the evaluation). Table 1 provides a breakdown of the documents included in this collection, TAC 2013 KBP Source Corpus (LDC2013E45). All of the newswire documents in the corpus were drawn from English Gigaword Fifth Edition (LDC2011T07), Chinese Gigaword Fifth Edition (LDC2011T13), and Spanish Gigaword Third Edition (LDC2011T12). All web documents in the package were drawn from various collections previously compiled for the GALE project. Discussion forums, newly added in 2013, were taken from a collection developed for the BOLT project in order to foster research on informal texts.

Language	Genre	Documents
English	Newswire	1,000,257
	Web Text	999,999
	Discussion Forums	99,063
Chinese	Newswire	2,000,256
	Web Text	815,886
	Discussion Forums	199,321
Spanish	Newswire	910,734

Table 1: 2013 Document Source Collection for Entity Linking and Regular, Sentiment, and Temporal Slot Filling Tasks (LDC2013E45)

The reference knowledge base (KB) (LDC2009E58) used in both the Entity Linking and Slot Filling tasks includes 818,741 nodes – articles drawn from an October 2008 dump of English Wikipedia. Each node corresponds to a unique entity corresponding to one of four types: person (PER), organization (ORG), geopolitical-entity (GPE), or unknown (UNK). All entries have semi-structured 'infoboxes', or tables of attributes pertaining to the subject entities. Some of the pages from the Wikipedia dump were not included in the KB because of ill-formatted infoboxes.

3. Training and Evaluation Data

As 2013 marked LDC's fifth year of supporting KBP evaluations, developers participating in this year's evaluations were able to receive a wealth of materials for training their systems. Including packages created in 2013, there are now 16 corpora of Entity Linking data, 16 for regular Slot Filling data, 5 Temporal Slot Filling packages, 5 Cold Start releases, and 3 collections of annotated data for Sentiment Slot Filling (see tables below for details).

Corpus Title (Dataset)	Type	LDC Catalog	Language	Size (Queries)
TAC 2009 KBP Gold Standard Entity Linking Entity Type List	Evaluation	LDC2009E86	English	567 GPE 627 PER 2710 ORG
TAC 2010 KBP Evaluation Entity Linking Gold Standard	Evaluation	LDC2010E82	English	749 GPE 741 PER 750 ORG
TAC 2010 KBP Training Entity Linking	Training	LDC2010E31	English	500 GPE 500 PER 500 ORG
TAC 2011 KBP Cross-lingual Training Entity Linking	Training	LDC2011E55	Chinese English	685 GPE 817 PER 660 ORG
TAC 2011 KBP English Evaluation Entity Linking Annotation v1.1	Evaluation	LDC2011R36	English	750 GPE 750 PER 750 ORG
TAC 2011 KBP Cross-lingual Evaluation Entity Linking Annotation V1.1	Evaluation	LDC2011R38	Chinese English	642 GPE 824 PER 710 ORG
TAC 2012 KBP Chinese Entity Linking Evaluation Annotations	Evaluation	LDC2012E103	Chinese English	605 GPE 699 PER 718 ORG
TAC 2012 KBP Chinese Entity Linking Web Training Queries and Annotations	Training	LDC2012E66	Chinese English	52 GPE 52 PER 54 ORG
TAC 2012 KBP English Entity Linking Evaluation Annotations	Evaluation	LDC2012E102	English	604 GPE 919 PER 706 ORG
TAC 2012 KBP Spanish Entity Linking Evaluation Annotations	Evaluation	LDC2012E101	Spanish English	858 GPE 669 PER 539 ORG
TAC 2012 KBP Spanish Entity Linking Training Queries and Annotations	Training	LDC2012E67	Spanish English	566 GPE 683 PER 601 ORG
TAC 2013 KBP English Entity Linking Evaluation Queries and Knowledge Base Links	Evaluation	LDC2013E90	English	803 GPE 686 PER 701 ORG
TAC 2013 KBP Chinese Entity Linking Evaluation Queries and Knowledge Base Links	Evaluation	LDC2013E96	Chinese English	714 GPE 706 PER 735 ORG
TAC 2013 KBP Spanish Entity Linking Evaluation Queries and Knowledge Base Links	Evaluation	LDC2013E97	Spanish English	660 GPE 695 PER 762 ORG

Table 2: Entity Linking Training and Evaluation Data

Corpus Title	Type	LDC Catalog	Language	Size
TAC KBP 2009 Evaluation Slot Filling List	Evaluation	LDC2009E65	English	53 Queries
TAC KBP 2009 Assessment Results	Evaluation	LDC2009E90	English	10,416 Assessments
TAC 2010 KBP Training Slot Filling Annotation	Training	LDC2010E18	English	50 Queries
TAC 2010 KBP Evaluation Slot Filling Annotation	Evaluation	LDC2010R11	English	100 Queries
TAC 2010 KBP Assessment Results	Evaluation	LDC2010E61	English	25,511 Assessments
TAC 2010 KBP Training Surprise Slot Filling Annotation	Training	LDC2010E52	English	32 Queries
TAC 2010 KBP Evaluation Surprise Slot Filling Annotation	Evaluation	LDC2010E52	English	40 Queries
TAC 2011 KBP English Training Regular Slot Filling Annotation	Training	LDC2011E48	English	48 Queries
TAC 2011 KBP English Evaluation Regular Slot Filling Annotation V1.2	Evaluation	LDC2011E89	English	100
TAC 2011 KBP English Regular Slot Filling Assessment Results V1.2	Evaluation	LDC2011E88	English	28,041 Assessments
TAC 2012 KBP English Regular Slot Filling Evaluation Annotations V1.1	Evaluation	LDC2012E91	English	80 Queries
TAC 2012 KBP English Regular Slot Filling Assessment Results V1.2	Evaluation	LDC2012E115	English	22,885 Assessments
TAC 2012 KBP Spanish Slot Filling Training Queries and Annotations V1.2	Training	LDC2012E68	Spanish English	50 Queries
TAC 2013 English Regular Slot Filling per:title Training Data	Training	LDC2013E60	English	1949 Assessments
TAC 2013 English Regular Slot Filling Evaluation Queries and Annotations	Evaluation	LDC2013E77	English	100 Queries
TAC 2013 English Regular Slot Filling Evaluation Assessment Results V1.1	Evaluation	LDC2013E91	English	27,655 Assessments

Table 3: Training and Evaluation Data for Regular Slot Filling

Corpus Title	Type	LDC Catalog	Language	Size
TAC 2011 KBP English Training Temporal Slot Filling Annotation	Training	LDC2011E49	English	50 Queries
TAC 2011 KBP English Evaluation Temporal Slot Filling Annotation	Evaluation	LDC2012E38	English	100 Queries
TAC 2013 KBP English Temporal Slot Filling Training Queries and Annotations	Training	LDC2013E82	English	7 Queries
TAC 2013 KBP English Temporal Slot Filling Evaluation Queries and Annotations	Evaluation	LDC2013E86	English	273 Queries
TAC 2013 KBP English Temporal Slot Filling Evaluation Assessment Results	Evaluation	LDC2013E99	English	4,376 Assessments

Table 4: Training and Evaluation Data for Temporal Slot Filling

Corpus Title	Type	LDC Catalog	Language	Size
TAC 2012 KBP Cold Start Queries V1.1	Evaluation	LDC2012E105	English	385
TAC 2012 KBP Cold Start Assessment Results	Evaluation	LDC2012E116	English	5015 Assessments
TAC 2012 KBP Cold Start Automated Queries Assessment Results	Evaluation	LDC2013E39	English	779 Assessments
TAC 2013 KBP English Cold Start Evaluation Queries and Annotations V1.1	Evaluation	LDC2013E87	English	326 Queries
TAC 2013 KBP English Cold Start Evaluation Assessment Results	Evaluation	LDC2013E101	English	6,755 Assessments

Table 5: Training and Evaluation Data for Cold Start

Corpus Title	Type	LDC Catalog	Language	Size
TAC 2013 KBP English Sentiment Slot Filling Training Queries and Annotations	Training	LDC2013E78	English	160 Queries
TAC 2013 KBP English Sentiment Slot Filling Evaluation Queries and Annotations	Evaluation	LDC2013E89	English	160 Queries
TAC 2013 KBP English Sentiment Slot Filling Evaluation Assessment Results	Evaluation	LDC2013E100	English	5,160 Assessments

Table 6: Training and Evaluation Data for Sentiment Slot Filling

4. Annotation & Assessment Procedures and Methodologies

4.1 Entity Linking

The overall goals of query selection for Entity Linking did not change in 2013. As in previous years, annotators sought to collect the most confusable named entity mentions they could find in the corpus. A query's confusability is measured both by the number of distinct entities in the set of queries that are referred to by its namestring (polysemy) as well as the number of distinct namestrings in the pool that refer to the entity (synonymy). For example, the namestring "Smith" would be highly confusable because one could likely find numerous instances of it being used in the corpus to refer to different entities. Additionally, entities with numerous nicknames and shortened or misspelled versions of their names in the corpus were targeted to increase synonymy in the query set.

Entity Linking queries were selected with the intention of representing as evenly as possible the three entity types (PERs, ORGs, and GPEs) and the statuses of NIL (not linked to the KB) and non-NIL. As was done in previous years, each set of Entity Linking queries strove for a source document genre ratio of 2/3 newswire to 1/3 web or informal documents. Lastly, for the cross-lingual versions of the task, although the majority of the queries were to be drawn from non-English documents, mentions in English documents of entities co-referential with other non-English queries were selected whenever possible.

To select queries, annotators searched the corpus, sometimes utilizing tagger output as a guide, and annotated any named entity mentions fitting the guidelines. Tagger output was used while searching for confusable namestrings (those that could refer to multiple entities) as searching a list of namestrings is more efficient than combing through whole documents. However, in searching for confusable entities (those who are referred to by multiple namestrings), annotators' creativity, world knowledge, and research skills were the most effective tools.

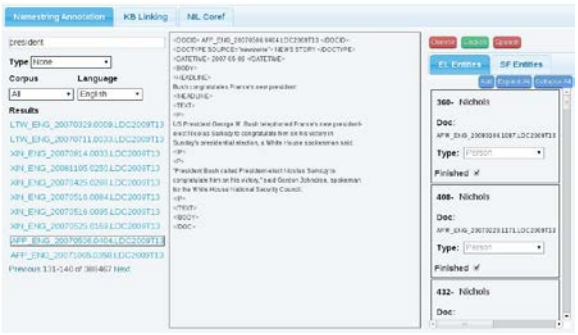

Figure 1: Namestring Annotation View of the Entity Selection Tool

There are three annotations phases to Entity Linking query development - namestring selection, knowledge base linking, and NIL coreference. However, while performing EL query development, LDC's online interface allows annotators to move back and forth between the three phases in order to more easily balance desired ratios of NIL and non-NIL queries and to break up, and thereby simplify, NIL coreference.

4.2 Regular Slot Filling, Temporal Slot Filling, Sentiment Slot Filling, and Cold Start

There is a great deal of similarity and overlap between the three versions of the Slot Filling task and Cold Start. All four tasks are made up of three generally separate processes – query development, annotation, and assessment. While there are certainly differences between the tasks, which will be discussed below, we will detail the three processes for each collectively, both to avoid redundancy and to highlight subtle differences.

4.2.1 Guidelines Updates

Building upon lessons learned in 2012 and planning discussions for the 2013 evaluations with TAC KBP coordinators, LDC made updates to some of the existing task guidelines, most notably the *TAC KBP Slot Descriptions* and *TAC KBP Slot Filling Assessment* guidelines.

The definition of *per:title* was altered such that the organization within which a title was held would henceforth be taken into account when determining whether or not a filler was redundant. For example, Mitt Romney has held three different "CEO" positions:

> CEO, Bain Capital (1984–2002)
> CEO, Bain & Company (1991–92)
> CEO, 2002 Winter Olympics Organizing Committee (1999–2002)

Even though the three titles are exactly the same, each of these responses would be placed into separate equivalence classes because the titles were held in distinct organizations.

Two other changes were made following observations of poor performance by both systems and annotators in previous evaluations. The first of these was the merging of *per:employee_of* and *per:member_of* into a single slot, *per:employee_or_member_of*. This change was made after noting the difficulty that both annotators and systems had in differentiating between the two slots in previous years. The second alteration was the reclassification of top-level governments of GPEs as GPEs themselves, rather than as ORGs as they had been classified in previous years and programs. This change proved particularly beneficial in Sentiment SF by allowing GPEs to be more readily included in relations. Given the two text extents below, examples such as the former are much more prevalent than the latter.

> The Palestinian government has denounced what it calls the Israeli army's 'current practice of shoot now and ask questions later.'
>
> We're kinda like David Hasselhoff; where we're big in Germany, but nobody else cares.

Justification, or minimum extents of provenance supporting the validity of a KBP relation, was also altered for the 2013 evaluations. Justification was added to Slot Filling in 2012 in an attempt to have systems and annotators highlight the sources of their assertions and, thereby, reduce the effort required for assessment. In 2012, justification was a single, minimal text extent proving the connection between the subject entity, via the selected slot, to the object entity, value, or string. In practice, the restriction to a single string often caused the provenance to include lengthy portions of unrelated text. As a result, justification was altered in 2013 to allow for multiple, discontiguous strings. For example, the following relation:

> <Harkat-ul-Mujahideen - *org:country_of_headquarters* - Pakistan>

could be maximally supported by the two following concise but possibly discontiguous text extents:

> the Islamabad headquarters of Harkat-ul-Mujahideen
>
> Islamabad, the capital city of Pakistan

4.2.2 Query Development

Much like Entity Linking, query development for all of the Slot Filling task varieties and Cold Start is driven by guided searches through the corpus. Unlike EL, however, initial searches usually focus on key words related to the KBP slots for the given task, rather than an entity name string. For example, annotators might search for "arrested" or "charged" to develop queries that will generate fillers for the *per:charges* slot. Once an initial 'seed' annotation such as the above is found, query developers search for other mentions of the connected entity or entities in the corpus to get a sense of how productive the query would be. Note, however, that while highly-productive queries are always desired, less productive queries that offer opportunities to fill under-utilized slots or slot types are also desired.

Task-specific selection criteria are also considered during the query selection processes. For regular Slot Filling, which uses single entities as queries, entity mentions must be non-

confusable. A candidate query was considered non-confusable if its namestring could be considered full (i.e. appropriate for use as the title of a Wikipedia page) and its referent could be easily identified by surrounding context. Additionally, the full set of Slot Filling queries was selected with the goal of representing approximately equally the three varieties of query types, namely, those that take named entities as fillers, those that take values as fillers (dates and numbers), and those that take strings as fillers. In previous Slot Filling evaluations, the restriction against returning fillers that were redundant with those already in the KB meant that non-NIL entities with fully fleshed out KB nodes were inappropriate as queries. However, as the redundancy restriction was lifted in the 2013 evaluation, such entities were acceptable as fillers (though still not preferred).

For the 2013 evaluations, the queries of Temporal Slot Filling, which requires performers to add temporal constraints to KBP relations based on textual evidence, were changed from the flagship task conducted in 2011, such that a full relation (e.g. "Bill Clinton" *per:spouse* "Hillary Clinton") acted as a query rather than just a single entity. This change was made primarily with the hope of generating greater participation in the evaluation by removing the need for a functional Slot Filling system in order to perform. A helpful byproduct of the decision to include slots within the queries was that representation of query types (based on their slots) was easy to control. As such, considerations for TSF query development were able to focus primarily on the richness and uniqueness of the temporal information that queries would generate. In 2011, many queries only generated WITHIN annotations, meaning that the specified relation held true on the publication date of a given source document. As such, for the 2013 queries, annotators ensured that the queries would allow for more interesting temporal information, such as indicators of beginnings and endings. For example, when searching for a potential *per:member_or_employee_of* query, annotators might search for terms such as "hired", "fired", or "quit".

Like Temporal Slot Filling, Sentiment Slot Filling in 2013 also included slots as part of the queries, allowing for easier control of equal slot representation. Accordingly, SSF query developers could focus on making queries highly productive, capable of generating edge-case or interesting fillers, or, ideally, both. An example of a more interesting response for a SSF query is "Carlson" for the query "Michael Vick" *per:neg-from* based on the provenance

> I think Michael Vick should have been executed for that", said Carlson

Correctly extracting a response from such a statement would be more challenging than from, say, "Carlson said he hated Michael Vick" due to the inference needed to catch the negative sentiment as well as the feeling's basis in an action. Although 2013 was the first year for Sentiment Slot Filling, query developers built sufficient challenge into the task so that the state of the art and, thereby, future directions for SSF, could be determined.

Cold Start query developers searched through the corpus selected specifically for the task and looked for entities richly connected to others via KBP slot relations. For example, given the following text extent:

> "Jane Doe is the president of the School of Arts and Sciences at the University of Pennsylvania"

annotators could create the following query:

"Jane Doe"
per:employee_of
"School of Arts and Sciences"
org:parents
"University of Pennsylvania"

Note that, while the example above only lists a single filler for each of the slots, there could potentially be multiple fillers at each "hop" level, all of which must be annotated and correctly connected to one another. This marks a notable difference between Cold Start and the other Slot Filling tasks, namely, full annotation for Cold Start queries is conducted at the time of query development rather than be treated as a later,

separate task. This difference is based largely in the fact that, while investigating fillers at the first hop level to determine what subsequent slot would be the most productive, query developers become so familiar with the elements of the query that it is simply more efficient to have them complete the annotation as well.

Validity decisions for Cold Start fillers are based on the same slot descriptions used for regular Slot Filling. However, in an attempt to increase connectivity between entities in the Cold Start corpus, inverse versions of all existing slots are also used. For example, for the existing slot *per:employee_or_member_of*, which captures organizations with which a person entity is affiliated as a member or employee, the inverse slot *org:employees_or_members* is used in order to also capture people who were affiliated with an organization entity.

4.2.3 Annotation

For each query in regular, Temporal, and Sentiment Slot Filling, annotators were given up to two hours to search the corpus and locate all valid fillers. Following the initial round of annotation, a quality control pass was conducted to flag any fillers that did not have adequate justification in the source document, or that might be at variance with the current guidelines. These flagged fillers were then adjudicated by senior annotators.

As was done in all previous regular Slot Filling evaluations, information from the Wikipedia infoboxes for query entities linked to the KB during entity selection was mapped to one or more of the TAC KBP slots. For example, if a given PER entity had "Philadelphia, PA" as its listed location of death in Wikipedia, that information would be separated into two filler strings ("Philadelphia" and "Pennsylvania") and mapped to the KBP slots *per:city_of_death* and *per:state_of_death*. Mappings were performed automatically and manually before results were reviewed and edited for consistency.

4.2.4 Assessment

Annotator training and testing was performed as a preliminary step for all Assessment tasks. After an initial training session and guidelines review, candidate assessors were required to complete an assessment screening kit, which contained 50 sample responses selected from past KBP evaluations. Assessors were required to assess every slot in the test kit and achieve 90% or higher accuracy for all slots. Those who passed the test went on to assess and coreference responses.

From an annotator's perspective, the Slot Filler assessment tasks are nearly identical except for some of the variations between the slots used. Fillers are marked as 'correct' if they are found to be supported in the reference document and in-line with the slot descriptions. Fillers are considered 'wrong' if they do not meet both of the conditions for correctness and 'inexact' if overly insufficient or extraneous text was selected for an otherwise correct response. The three main components of justification – subject mentions, object mentions, and the full predicates – are also assessed as correct, wrong, and inexact though predicates can be more specifically 'inexact-short' or 'inexact-long'. Assessors also had the option of ignoring full responses if their justification strings were considered too long to merit review.

After assessment was completed, quality control was performed on the data using a procedure similar to that described above for slot filling annotation, in which annotators reviewed the work of their peers and flagged potentially problematic assessments for additional review. As with the Slot Filling quality control procedure, this process improved assessment results while also indicating potential improvements in the guidelines and areas in which some annotators required more training.

5. Conclusion

This paper discussed procedures and methodologies for annotation and assessment for KBP 2013, particularly elaborating on procedures and methodologies for query

selection, annotation, and assessment. LDC support of KBP in 2013 included source corpus expansion; revisions to the entity selection processes for both the Entity Linking and Slot Filling tasks in order to support coordinator requests for more challenging and diverse queries; revision of the annotation process and data collected for Slot Filling; expansion of cross-lingual data with the addition of Spanish Entity Linking; as well as the addition of a two new tasks – Temporal Slot Filling and Sentiment Slot Filling – which brought the total number of tasks supported in 2013 to seven, one more than in 2012. Future work will include further refinement of the changes made to tasks this year and development of new tasks. The resources described in this paper are slated for publication in the LDC Catalog, in order to make the corpora available to the wider research community. Other resources such as KBP system descriptions and site papers will be published on the NIST TAC website.

References

Hoa T. Dang, Jimmy Lin, and Diane Kelly. 2006. Overview of the TREC 2006 Question Answering Track. In Proceedings of TREC 2006.

George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel. 2004. Automatic Content Extraction (ACE) program - task definitions and performance measures. In Proceedings of the Fourth International Language Resources and Evaluation Conference.

Ralph Grishman, David Westbrook, and Adam Meyers. 2005. NYU's English ACE 2005 System Description. In Proceedings of the ACE 2005 Evaluation/PI Workshop.

Paul McNamee, Hoa T. Dang, Heather Simpson, Patrick Schone, and Stephanie M. Strassel. 2010. An Evaluation of Technologies for Knowledge Base Population. In Proceedings of the Seventh International Language Resources and Evaluation Conference.

Heather Simpson, Stephanie Strassel, Robert Parker, and Paul McNamee. 2010. Wikipedia and the Web of Confusable Entities: Experience from Entity Linking Query Creation for TAC 2009 Knowledge Base Population. In Proceedings of LREC.