Task Description for
Knowledge-Base Population at TAC 2010

Version of: 06/14/2010

1 Introduction

The main goal of the Knowledge Base Population (KBP) track at TAC 2010 is to promote research in and to evaluate the ability of automated systems to discover information about named entities and to incorporate this information in a knowledge source. For the evaluation an initial (or reference) knowledge base will be provided along with a document collection that systems are to use to learn from. Attributes (a.k.a., "slots") derived from Wikipedia infoboxes will be used to create the reference knowledge base. There will be two related tasks: Entity Linking, where names must be aligned to entities in the KB, and Slot Filling, which involves mining information about entities from text. Slot Filling can be viewed as more traditional Information Extraction, or alternatively, as a Question Answering (QA) task, where the questions are static but the targets change. Groups may participate in either, or both tasks, though participation in both is encouraged.

Compared to previous information extraction evaluations such as MUC and ACE, KBP involves the following new research topics:

- Extraction at large scale (e.g. 1 million documents) ;
- Using a representative collection (not selected for relevance);
- Cross-document entity resolution (extending the limited effort in ACE);
- Linking the facts in text to knowledge base;
- Offering the possibility of distant (and noisy) supervision through Infoboxes;
- Rapid adaptation to new relations.

This is a follow-on to the KBP evaluation at TAC 2009, with a generally similar structure, training data for slot filling, and more preparation time for system development. Among the significant changes planned for the 2010 evaluation are to:

- Remove the task of linking of slot fills with the knowledge base;
- Remove GPEs from the slot filling task;
- Provide optional entity linking task without background wikipedia texts;
- Provide optional surprise slot filling tasks;
- Provide a simpler and more uniform scoring metric for slot filling;
- Emphasize genre diversity for entity linking, add significant amount of web data into the source collection.

The tasks will be structured by having participants process a list of target entities. For the Entity Linking task the list will contain entity types of Person (PER), Organization (ORG), and Geo-Political Entity (GPE). As in the ACE evaluation, GPEs include inhabited locations with a government such as cities and countries. For the Slot Filling task the list will only contain PER and ORG entities.

2 Knowledge Base

Wikipedia infoboxes will be the basis for the reference knowledge base; however exact compliance with Wikipedia is not intended. The KB will be derived from the set of entities from Wikipedia that have infoboxes, and it will contain hundreds of thousands of nodes. Each entity in the knowledge base (sometimes called a node) will include the following:

- a name string
- an assigned entity type of PER, ORG, GPE, or UKN (unknown)
- a KB node ID (a unique identifier, like "E101")
- a set of 'raw' (Wikipedia) slot names and values
- some disambiguating text (i.e., text from the Wikipedia page)

For PER and ORG entities there is a set of desired attributes ("slots"). Guidelines for each slot are available at: http://projects.ldc.upenn.edu/kbp/. The guidelines specify whether the slots are single-valued (e.g., per:date_of_birth) or list-valued (e.g., per:employee_of, per:children). Official names for each slot are given in Table 1.

Person	**Organization**
per:alternate_names	org:alternate_names
per:date_of_birth	org:political/religious_affiliation
per:age	org:top_members/employees
per:country_of_birth	org:number_of_employees/members
per:stateorprovince_of_birth	org:members
per:city_of_birth	org:member_of
per:origin	org:subsidiaries
per:date_of_death	org:parents
per:country_of_death	org:founded_by
per:stateorprovince_of_death	org:founded
per:city_of_death	org:dissolved
per:cause_of_death	org:country_of_headquarters
per:countries_of_residence	org:stateorprovince_of_headquarters
per:stateorprovinces_of_residence	org:city_of_headquarters
per:cities_of_residence	org:shareholders
per:schools_attended	org:website
per:title	
per:member_of	
per:employee_of	
per:religion	
per:spouse	
per:children	
per:parents	
per:siblings	
per:other_family	
per:charges	

Table 1. Slot Names for the Two Generic Entity Types

The 'raw' slot names and the values in the reference KB are based on an October 2008 Wikipedia snapshot. Wikipedia infoboxes are not an ideal knowledge representation; one key disadvantage is a lack of inheritance (and therefore, consistency). Wikipedia infoboxes also tend to focus on presentation on a Wikipedia page instead of abstract representation. As an example, consider the table below. The infobox for each of these organizations contains a slot related to the date it was created; however, the name of the slot varies.

Organization	Infobox Class	Slot Name	Slot Value
Bill/Melinda Gates Foundation	Non-Profit	founded_date	1994
Cornell	University	established	1865
FDA	Government_Agency	formed	1906
NASA	Government_Agency	formed	July 29, 1958
Washington Redskins	NFL_Team	founded	1932

Table 2. Examples of Wikipedia Infobox Slot Name and Slot Value Variability

To facilitate use of the reference KB a mapping from raw Wikipedia infobox slot-names to generic slots will be provided. For example, Wikipedia infobox slots named *established* or *creation_date* may contain a value for *org:founded*. These mappings should not be viewed as precise synonyms. Sometimes Wikipedia slots contain multiple pieces of information that belong in separate generic slots. For example, some Wikipedia infoboxes contain a slot named '*born'* that has values for both *per:date_of_birth* and *per:place_of_birth*. Also, the 'semantics' of a Wikipedia infobox field will not always correspond with the generic slot guidelines. For example, the children slot in Wikipedia's Person infobox sometimes gives a number to indicate that the entity has that number of children; however, only names of children are acceptable for the generic slot *per:children*.

3 Entity Linking

3.1 General Task

The Entity Linking task is to determine for each query, which knowledge base entity is being referred to, or if the entity is not present in the reference KB. A query will consist of a name-string and a document-id in the test collection. Each name-string will occur in the associated document in the test collection. The purpose of the associated document is to provide context that might be useful for disambiguating the name-string. Entities will generally occur in multiple queries using different name variants and/or different docids. It is also expected that some entities will share confusable names (e.g., *George Washington* could refer to the president, the university, or the jazz musician; *Washington* could refer to a city, state, or person).

Queries should be processed independently from one another. This is not meant to prohibit parallel computation on clusters, but only to require that systems may not leverage knowledge about the set of queries. Examination of multiple queries to make a better decision about a single query is expressly prohibited.

For each query a KB node-id must be returned. For entities that have no corresponding node in the reference KB a response of NIL should be returned. System responses will be evaluated based on the correctness of linkages to KB nodes. Please note that for KBP2010 an entity linking system is not required to cluster non-KB entities together. In order to make the entity linking task more realistic, web data is added to the source collection.

3.2 Optional Task: Entity Linking without Wikipedia pages

For the primary task, the system may consult the text from the Wikipedia pages associated with the knowledge base nodes. There will be also an optional task in which the systems should do linking without reference to these texts. – using only the slot values; this corresponds to the task of updating a knowledge base with no 'backing' text.

3.3 Query and Output Formats

Queries will be formatted in XML. Here is an example of the formatting:

<query id="EL1"><name>John Doe</name><docid>SUN-009</docid></query>
<query id="EL2"><name>John Doe</name><docid>NYT-005</docid></query>
<query id="EL3"><name>Johnny Doe</name><docid>CNN-001</docid></query>

System output files should have a single response per line, and consist of two space-separated columns. The first column is the query id from the input file and the second column is either a knowledge base id, or NIL. For example:

EL1 E101
EL2 NIL
EL3 E5871

Thus, the name-string from query "EL1" is believed to refer to the entity in knowledge base node E101, and the entity in query "EL2" is believed to be absent from the reference KB. The DTD for entity linking is available at: http://projects.ldc.upenn.edu/kbp/dtd/kbpentlink.dtd.

3.4 Scoring Metric

The entity linking task is similar to cross-document co-reference; however, here the problem requires alignment to a knowledge base, not clustering of entities. The official evaluation measure will be micro-averaged accuracy. An example is given below.

Query	System Assignment	Correct Assignment	Correct?
EL1	E101	E101	Yes
EL2	NIL	E101	No
EL3	E5871	E5871	Yes
EL4	E101	E101	Yes
EL5	NIL	NIL	Yes

Table 3. Entity Linking Scoring Example

Here 4 out of 5 responses are correct, so the micro-averaged accuracy is 0.80. Micro-averaged accuracy is the sole official measure. A script which computes micro-averaged accuracy, along with macro-averages across entities, is available at: http://nlp.cs.qc.cuny.edu/kbp/2010/scoring.html.

4 Slot Filling

4.1 General Task

The Slot Filling task involves learning a pre-defined set of relationships and attributes for target entities based on the documents in the test collection. A query in the Slot Filling task will contain a name-string, docid, entity-type, node-id, an optional list of slots to ignore. For example: [Paul Newman, ABC-20080611-9372, PER, E2317, per:employee-of] might be a query for actor Paul Newman. The node id that is provided will refer to a node representing the entity in the KB. For targets for which no node exists in the KB, the node-id will begin with "NIL", e.g., "NIL102". As in the entity linking task the provided docid is intended to give context for the entity. The list of slots to fill will indicate that responses should be returned for these slots.

Systems must process the target entities (i.e., each query) independently from one another. For each slot value returned, systems must also return a single docid from the test collection that supports the value returned for the given entity and slot.

Slots can be of one of two types: single-valued slots that admit only a single value (e.g., per*:date_of_birth*) and list-valued slots that can accept more than one value (e.g., per*:employee_of*). In some cases multiple correct and supportable values may exist in the corpus for a single-valued slot. For example, there may be distinct values for *per:age, per:religion.* or *org:website.* In such cases, any correct and supported response is sufficient.

Systems are not expected to correct or modify values from the reference KB, but only to add information. Therefore no information is sought for single-valued slots that already have a value in the KB node.

Redundant information should not be returned; only novel information is of interest. However, if an attribute has a value in the initial KB, the slot should not necessarily be ignored entirely. For example, if the Wikipedia alma-mater slot for investor Warren Buffet looked like:

<fact name="alma-mater">University of Nebraska</fact>

but a document is found which notes that he graduated from Columbia University, then this should be returned as a value for *per:schools_attended* (because *per:schools_attended* allows multiple values). However, University of Nebraska should not be returned, because it is redundant with what is already in the initial KB. Similarly, if multiple equivalent values occur in the test collection, the value should be returned only once, with any one of the supporting docids (i.e., Columbia University should be returned only once even if there are multiple documents that support it as a value for *per:schools_attended).*

It is not required that slot values be a contiguous span of text from the supporting document. However, values directly extracted from the text or responses that remain as close as possible to the source text are preferred. Returning "10/31/1951" when a document states "Oct. 31, 1957" is reasonable (although such normalization is not required), but returning "Halloween 1951" or "the fifth Wednesday in Oct. 1951" from the same document is not likely to be accepted because these would be very unusual forms for that information in a knowledge base. Timestamps on documents can be used in determining dates; thus, if a document refers to someone dying "July 31st last year", a year could be returned in *per:date_of_death.*

4.2 Optional Task: Surprise Slot Filling

A central track of IE research is the issue of *portability* – how can an IE system rapidly and automatically (semi-automatically) move to new domains and new slot (relation/event) types. In KBP 2010 we are organizing a new "surprise" slot filling task. The adaptation could be fully automatic (based on examples) or involve limited human intervention. One of the novel features of this year's KBP Slot Filling task will be the Surprise Slot task. Our goal in including this task is to showcase some of the R&D, involving semi-supervised learning, active learning, and distant learning, which allows the rapid retargeting of extraction systems to new tasks.

The Surprise task will involve a small number of slots. We are currently planning on two slots for person entities and two slots for organization entities. The slots may be either single-valued or list-valued slots. The Surprise task will use the same text corpus as the main slot filling task.

We will provide for each slot a one paragraph description. We will also provide a small set of training entities with annotations for these four slots. The size of training and evaluation data for the surprise task is described in section 5.2. The annotations will be provided in the same (11-field) format as training data.

The surprise evaluation data will be in the same format as the test data for the main slot filling task. Note that there is no necessary connection between the evaluation entities for the main task and the Surprise task. For the entities in the Surprise task, systems should *only* provide fills for the Surprise slots. The evaluation schedule is shown in Table 8.

The results will be presented in a two-dimensional plot reflecting task performance (the same F measure as for the main task) and elapsed time, in hours, from the time the training data is downloaded until the evaluation results are uploaded. Ideally, this would show some trade-off between faster systems (e.g., fully automated semi-supervised training which can do the task in

one hour) and higher-performance systems using more extensive human input. No preference will be defined regarding the trade-off of speed and performance -- we are not looking for a unique winner. Each participating site will be asked to self-report some additional numbers: the breakdown into training and run time, and the size of the team actively participating in the Surprise task.

As with the main evaluation task, sites will be permitted to submit up to three alternative runs for the Surprise task.

4.3 Query and Output Formats

Slot filling queries will be formatted in XML. Here is an example of the formatting:

```
<query id="SF1">
  <name>John Doe</name>
  <docid>SUN-009</docid>
  <enttype>PER</enttype>
  <nodeid>E101</nodeid>
  <ignore>per:date_of_birth per:place_of_birth per:religion</ignore>
</query>



<query id="SF2">
  <name>ACME Widget Corp</name>
  <docid>NYT-006</docid>
  <enttype>ORG</enttype>
  <nodeid>NIL102</nodeid>
</query>
```

System output files should contain at least one *response* for each query-id/slot combination, except that no response should be returned for slots listed in the <ignore> field. A response consists of a single line, with a separate line for each slot value. Lines should have the following space-separated columns:

Column 1: query id
Column 2: the slot name
Column 3: a unique run id for the submission
Column 4: NIL, if the system believes no information is learnable for this slot. Or, a single docid which supports the slot value
Column 5: a slot value

Except for the last column containing the slot value, the columns cannot contain whitespace characters. When no novel information is believed to be learnable for a slot, Column 4 should be NIL and Column 5 should be left empty.

For each query, the output file should contain exactly one line for each single-valued slot. For list-valued slots, the output file should contain a separate line for each list member. A response like “Tropicana Products and Frito-Lay” would be considered inexact for *org:subsidiaries*.

The file should be sorted by slot filling query id and multiple responses for the same list slot should be contiguous. But it is not necessary to order the slots for individual queries. For example:

```
SF1 per:spouse uva1x NBC-3218 Jane Smith
SF1 per:spouse uva1x CNN-387 Mary Doe
```

SF1 per:date_of_death uva1x ABC-007 April 23, 2008
SF1 per:children uva1x CNN-387 Bobby Doe
SF1 per:employee_of uva1x NIL
SF1 per:schools_attended uva1x SUN-3321 Cornell
SF1 per:schools_attended uva1x SUN-3321 Harvard Law School
SF1 per:schools_attended uva1x SUN-3321 NYU Law School
...
SF2 org:headquarters uva1x NYT-001 Cleveland, Ohio
SF2 org:alternate_names uva1x NYT-701 Widgets-R-Us
SF2 org:founded_by uva1x ABC-119 John "Hammer" Smithson
SF2 org:website uva1x NIL
SF2 org:parents uva1x NIL
SF2 org:subsidiaries uva1x NIL

The DTD for slot filling is available at: http://projects.ldc.upenn.edu/kbp/dtd/kbpslotfill.dtd

4.4 Evaluation

In contrast to the 2009 evaluation, a uniform scoring metric for all slots is planned, based on traditional measures of recall, precision, and F-measure, computed from counts of correct, missing, and spurious responses. A non-NIL response is correct if it matches a verified non-NIL entry in the key (the human assessment file); other non-NIL responses are spurious. A NIL response where the key has a verified non-NIL response is considered missing. NIL system responses matching verified NIL entries in the key *are not counted*. For single-valued slots only a single system response will be accepted. For list-valued slots, the verified non-NIL responses will be grouped into equivalence classes. Multiple responses to a query must come from disjoint classes to be counted as correct; other responses are counted as spurious.

5 Data

5.1 Data Selection

KBP 2010 training and evaluation data will be selected using a careful targeted process. We will provide training and evaluation corpora for both tasks. For entity linking, we will aim to achieve a balance between entity disambiguation and alias detection. For slot filling, we will select for training data entities with many filled slots, or which have information for slots which are under-represented, or slots with general poor performance. In addition to newswire data, we will also provide some web data to stimulate genre-specific research.

5.2 Training and Evaluation Corpus

Registered participants who did not participate in 2009 will initially receive the 2009 entity linking training data and the human assessments from the 2009 evaluation for both tasks. LDC will augment the current slot-filling answer keys with alternative answers found through time-limited human search. Answer keys for slot filling for some additional entities (persons and organizations) will also be provided (see the detailed schedule in Table 8).

The primary text corpus will be a collection of approximately 1 million news articles already used in 2009; LDC will expand this with a significant amount of selected web data (blogs etc.). The reference knowledge base includes hundreds of thousands of entities based on articles from an October 2008 dump of English Wikipedia which includes 818,741 nodes.

In KBP 2010, each participant is asked to annotate slot fills for several entities and share the results with the other sites. There will be some overlap of entities among sites to check inter-annotator agreement. LDC will share their annotation tool if and when that proves feasible.

The training and evaluation corpora statistics including source collection and target entity lists

are listed in the following tables. To verify data format integrity, the DTDs are available for the knowledge base at: http://projects.ldc.upenn.edu/kbp/dtd/knowledge_base.dtd, and for the source collection at; http://projects.ldc.upenn.edu/kbp/dtd/ace_source_sgml.v1.0.5.dtd.

Genre	Approximate Size (documents)
Newswire	1 million
Web data	300,000

Table 4. Source Collection Corpus

Corpus	Genre/Source	Size (entity mentions)		
		Person	Organization	GPE
Training	2009 Training	627	2710	567
	2010 Web data	500	500	500
Evaluation	Newswire	500	500	500
	Web data	250	250	250

Table 5. Entity Linking Corpora

Corpus	Task	Source	Size (entities)	
			Person	Organization
Training	Regular Task	2009 Evaluation	17	31
		2010 Participants	25	25
		2010 LDC	25	25
	Surprise Task	2010 LDC	16	16
Evaluation	Regular Task	Mixture	50	50
	Surprise Task	Mixture	20	20

Table 6. Slot Filling Corpora

6 External Resource Restrictions and Sharing

6.1 External Resource Restrictions

For 2010, participants will be asked to make at least one run subject to certain resource constraints, primarily that the run be made as a 'closed' system … one which does not access the Web during the evaluation period. Sites may also submit an additional run with access the Web. This will provide a better understanding of the impact of external resources.

Further rules for both of the primary runs and additional runs are listed in Table 7.

Specific Rules	Specific Examples
Allowed	Using a Wikipedia derived resource to (manually or automatically) create training data
	Compiling lists of name variation based on hyperlinks and redirects before evaluation
	Using a Wikipedia derived resource before evaluation to create a KB of world knowledge which can be used to check the correctness of facts
	Preprocess/annotate a large text corpus before the evaluation to check the correctness of facts or aliases

Not Allowed	Using Wikipedia infoboxes to directly fill slots
	Editing Wikipedia pages for target entities, either during, or after the evaluation

Table 7. Rules of Using External Resources

6.2 Resource Sharing

In order to support groups which intend to focus on part of the tasks, the participants are encouraged to share the external resources that they prepared before the evaluation. The possible resources may include intermediate results, entity annotations, parsing/SRL/IE annotated Wikipedia corpus, topic model features for entity linking, patterns for slot filling, etc. The sharing process can be informal (among participants) or more formal (through a central repository built by the coordinators). Please email the coordinators in order to access the central site.

7 Submissions and Schedule

7.1 Submissions

In KBP 2010 each participating site is allowed to select an evaluation window between June 4 and July 26. A site will have one week after downloading the data to return their results. Up to three alternative system runs may be submitted by each team for each of the two tasks. Systems should not be modified once queries are downloaded. Details about submission procedures will be communicated to the track mailing list. A script will be made available to ensure that submission files comply with the prescribed format.

7.2 Schedule

A schedule for KBP 2010 is presented in Table 8.

Date	Event
01/15	Preliminary task definition available
01/15	Slot filling scorer available
01/25	Preliminary annotation guideline available
02/03	Revised annotation guideline available
02/05	Revised task definition available
02/17	Registration site available
02/17	Entities for participant annotation available
03/29	Release of all 2010 source data (old 2009 data + 2010 web data)
03/31	Final annotation guideline available
04/05	Slot filling training corpora based on 09eval entities available
04/30	First release of entity linking training data (web texts)
05/01	Slot filling training corpora from participants available
05/12	Final release of entity linking training data (web texts)
05/14	Slot filling training corpora based on 2010 entities available
05/10	Dry run entities available [2 entities]
05/12	Dry run results due
05/18	Dry run assessment results available
05/21	Registration deadline
06/11	Entity linking and Regular slot filling task evaluation data (target query lists) available
06/04 –07/26	Participant to select a one-week evaluation window
07/26	Surprise training data and test entities available (10AM EST)

07/30	Surprise slot filling evaluation results due
09/15	Assessments for all tasks available
10/27	System description due
11/15-11/16	TAC 2010 workshop (NIST)

Table 8. KBP 2010 Schedule

8 Mailing List and Website

The KBP 2010 website is http://nlp.cs.qc.cuny.edu/kbp/2010 . Please post any questions and comments to the list tac-kbp@nist.gov. Information about subscribing to the list is available at: http://nlp.cs.qc.cuny.edu/kbp/2010/mailing.html.