Proposed Task Description for Knowledge-Base Population at TAC 2013

English Slot Filling – Regular and Temporal

Version 1.1 of May 24th, 2013

1 Changes

- 1.0 – Initial release
- 1.1 – Added information about the "TAC 2013 KBP Source Corpus" (Catalog ID LDC2013E45) to Section 5.1.

2 Introduction

The main goal of the Knowledge Base Population (KBP) track at TAC 2013 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 source document collection from which systems are to learn. Attributes (a.k.a., "slots") derived from Wikipedia infoboxes will be used to create the reference knowledge base. This document focuses only on the English Slot Filling (ESF) task, 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. For the other tasks part of KBP 2013, please visit the KBP web page: http://www.nist.gov/tac/2013/KBP/.

Compared to the KBP slot filling evaluation at TAC 2012, we aim to achieve the following new research goals:

- Provide provenance and justification texts that are, at the same time, concise and informative (see Section 2.2 for details). These provenance and justification texts allow the assessor to quickly see the points in the document where the fact is attested, and, in the future, will provide further training data for systems attempting to learn contextual patterns for slots.
- Revive the research on temporal slot filling. This year offers a dedicated track on temporal slot filling, which is a continuation of the 2011 pilot. To lower the barrier of entry for this task, this year's input for the temporal slot filling task will provide *both* the entity and the slot filler to be analyzed (akin to the diagnostic task in 2011).
- We have extended the document collections for the English slot filling tasks with data from discussion forums, which should foment research on information extraction from less formal texts.

3 English Regular Slot Filling

The goal of Slot Filling is to collect from the corpus information regarding certain attributes of an entity, which may be a person or some type of organization. Guidelines for each of the slots will be available at: http://surdeanu.info/kbp2013/def.php. 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 KBP 2013 slot are given in Table 1. The guidelines for KBP 2013 slots are close to the guidelines for KBP 2012, with a couple of significant changes:

1. The definition of the per:title slot changed significantly. This year, titles that report positions at different organizations are reported separately (see Section 2.3 and the slot annotation guidelines for details).
2. The per:employee_of and per:member_of slots were merged into a single slot, per_employee_or_member,_of due to their similarity. The semantics of the new slot also changed somewhat (see Section 2.3 and the slot annotation guidelines for details). We will attempt not to change the remaining the slot types.[1] See Table 1 for the complete list of slots addressed in this year's evaluation.
3. This year, entities mentioned in document meta data can be used as input for the slot fillings tasks or fillers to be extracted by systems. For example, systems should consider as slot filler candidates the post authors, which are recorded in the meta data of discussion forum documents.

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:date_founded
per:city_of_death	org:date_dissolved
per:cause_of_death	org:country_of_headquarters
per:countries_of_residence	org:stateorprovince_of_headquarters
per:statesorprovinces_of_residence	org:city_of_headquarters
per:cities_of_residence	org:shareholders
per:schools_attended	org:website
per:title	
per:employee_or_member_of	
per:religion	
per:spouse	
per:children	
per:parents	
per:siblings	
per:other_family	
per:charges	

Table 1. KBP2013 Slot Names for the Two Generic Entity Types

3.1 Input Format

Each query in the Slot Filling task consists of the name of the entity, its type (person or organization), a document (from the corpus) in which the name appears (to disambiguate the

[1] We are still reviewing the annotation process from last year's evaluation. This might result in changes to slot definitions. We will however try to minimize differences from last year's definition.

query in case there are multiple entities with the same name), the start and end offsets of the name as it appears in the document, its node ID (if the entity appears in the knowledge base), and the attributes which need not be filled. An example query is:

<query id="SF_002">
<name>PhillyInquirer</name>
<docid>eng-NG-31-141808-9966244</docid>
<beg>757</beg>
<end>770</end>
<enttype>ORG</enttype>
<nodeid>E0312533</nodeid>
<ignore>org:city_of_headquarters org:country_of_headquarters org:date_founded org:number_of_employees_members org:stateorprovince_of_headquarters org:website</ignore>
</query>

Note that unlike previous years, this year's task removes the non-redundancy requirement with the KB for filler. That is, slot fillers that are already filled in the reference database *must* be reported as well. We hope this simplifies system development, as developers do not have to implement a redundancy component. However, slots listed in the <ignore> field must be removed from the submitted output (similar to previous years). For example, for the above query, systems should not extract fillers for org:city_of_headquarters, org:country_of_headquarters, and org:date_founded.

Along with each slot filler, the system must provide a confidence score, provenance for both filler and query entity, and justification for the extraction. See Section 2.2 for details on the output format. If the corpus does not provide any information for a given attribute, the system should generate a NIL response (and no provenance or confidence score).

For each attribute we indicate the type of fill and whether the fill must be (at most) a single value or can be a list of values. For list-valued slots, fillers returned for the same entity and slot must refer to distinct individuals. It is not sufficient that the strings be distinct; for example, if the system finds both "William Jefferson Clinton" and "Bill Clinton" as fillers for the same entity and slot, it should return only one of those fillers (the other would be considered redundant and reduce system precision).

3.2 Output Format

System output files should be in UTF-8 and 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 tab-separated columns:
Column 1: query id
Column 2: 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 that justifies the relation between the query entity and the slot filler
Column 5: a slot filler (possibly normalized, e.g., for dates)
Column 6: start-end offsets for representative mentions used to extract/normalize filler
Column 7: start-end offsets for representative mentions used to extract/normalize query entity
Column 8: start-end offsets of clause(s)/sentence(s) in justification
Column 9: confidence score

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. When no novel information is believed to be learnable for a slot, Column 4 should be NIL and Columns 5-9 should be left empty. Column 5 (if present) contains the string representing the slot filler; the string should be extracted from the document in Column 4, except that any embedded tabs or newline characters should be converted to a space character and dates must be normalized. Systems have to normalize document text strings to standardized month, day, and/or year values, following the TIMEX2 format of yyyy-mm-dd (e.g., document text "New Year's Day 1985" would be normalized as "1985-01-01"). If a full date cannot be inferred using document text and metadata, partial date normalizations are allowed using "X" for the missing information. For example:

- May 4th" would be normalized as "XXXX-05-04";
- "1985" would be normalized as "1985-XX-XX";
- "the early 1900s" would be normalized as "19XX-XX-XX" (note that there is no aspect of the normalization that captures the "early" part of the filler).

See the the assessment guidelines document (version V3.2) for more details on the normalization requirements.

Columns 6 through 8 rely on offsets in documents. Note that each document is represented as a UTF-8 character array and begins with the "<DOC>" tag, where the "<" character has index 0 for the document. Thus, offsets are counted *before* XML tags are removed. In general, start offsets in these columns must be the index of the first character in the corresponding string, and end offsets must be the index of the last character of the string (therefore, the length of the corresponding mention string is endoffset – startoffset + 1).

Provenance: Columns 6 through 8 must contain the provenance of the slot filler string in the document, the provenance of the query entity in the document, and the justification for the extraction. To account for the fact that systems may use coreference resolution and date normalization to extract or match the slot filler and entity, Columns 6 and 7 must contain at least one mention and may contain the offsets of up to two relevant mentions, i.e., up to two pairs of start/end offsets. It is mandatory to include the start/end offsets of the filler/entity mentions used for extraction, which must be contained in the justification reported in Column 8. If additional mentions (i.e., document date or mentions from a coreference chain) were used for the normalization of the corresponding filler or entity, the offsets of what the system considered the most representative mention, i.e., the fullest or most informative name string in the document, must be included. For example, a date in Column 5 that is computed from the document date and the string "yesterday" should provide the offsets for both "yesterday" and the document date in Columns 6.

Offset formatting: Start and end offsets should be separated by dash ("-") and pairs of start/end offsets for different mentions should be separated by comma (","). For example, for the above query, if "yesterday" appears at offset 200 in the document and the document date appears at offset 20, then a valid entry for Column 6 in this case would be: 200-208,20-32 (assuming the endoffset for the document date is 32).

A more complicated example involves coreference resolution for both slot filler and entity. For example, consider the query per:spouse of "Michelle Obama" and the text:

> *Michelle Obama started her career as a corporate lawyer specializing in marketing and intellectual property. Michelle met Barack Obama when she*

was employed as a corporate attorney with the law firm Sidley Austin. She married him in 1992.

Column 8 must contain the location of a minimal number of clauses or sentences that provides justification for the extraction. *This column must contain at least one clause (or sentence) and at most two sentences. If two sentences are reported, they may be discontiguous.* Similar to Columns 6 and 7, start and end offsets must be separated by dash, and, if two clauses are reported, the start/end offset pairs must be separated by comma.
For example, assuming a system that performs extraction only from single sentences, Column 8 will contain the startoffset and endoffset of the sentence: "She married him in 1992." Column 6 must contain the offsets of the two mentions that were used to extract and normalize the slot, where one of the mentions must come from the justification text reported in Column 8. That is, for the above query and system, Column 6 must contain the offsets for both "him" and "Barack Obama". Similarly, Column 7 must contain the offsets of the mentions that were used to extract and normalize the entity. In the above example, Column 7 must contain the offsets for "She" and "Michelle Obama".

Note that whether or not Column 6 (or Column 7) requires multiple start/end offset pairs, will depend on the extent selected for Column 8. For example, a system that performs cross-sentence extraction might report in Column 8 the offsets of the text: "Michelle met Barack Obama when she was employed as a corporate attorney with the law firm Sidley Austin. She married him in 1992." In this situation, the slot filler "Barack Obama" is fully disambiguated in the justification text, so this system should report just the start/end offsets for "Barack Obama" in Column 6. However, the query entity is not fully disambiguated in the justification text. Thus, this system will have to report the start/end offsets for "Michelle" or "She" (whichever was used for extraction) and "Michele Obama".

Note that it is valid to report multiple clauses or sentences in Column 8 (either as a sequence of start/end offsets or as a single start/end offset pair for contiguous texts) *only* if the underlying system used the corresponding text during extraction. It is illegal to report the entire document or, in general, text that is not relevant for extraction in Column 8. In general, justification text that is too verbose, too small, or wrong will be considered incorrect.

A human assessor will judge the correctness of the (possibly normalized) slot filler string, and correctness of the offsets for the slot filler, query entity, and justification. We will report two different scores for this task: (a) ignoring the offsets, and (b) scoring the offsets, i.e., a slot filler will be considered correct only if the offsets in Column 6 through 8 are also correct.

Confidence Scores: To promote research into probabilistic knowledge bases and confidence estimation, each non-NIL response must have an associated confidence score. Confidence scores will not be used for any official TAC 2013 measure. However, the scoring system may produce additional measures based on confidence scores. For these measures, confidence scores will be used to induce a total order over the responses being evaluated; when two scores are equal, the response appearing earlier in the submission file will be considered to have a higher confidence score for the purposes of ranking. A confidence score must be a positive real number between 0.0 (representing the lowest confidence) and 1.0 (inclusive, representing the highest confidence), and must include a decimal point (no commas, please) to clearly distinguish it from a document offset. In 2013, confidence scores may not be used to qualify two incompatible fills for a single slot; submitter systems must decide amongst such possibilities and submit only one. For example, if the system believes that Bart's only sibling is Lisa with confidence 0.7 and Milhouse

with confidence 0.3, it should submit only one of these possibilities. If both are submitted, it will be interpreted as Bart having two siblings.

NIST reserves the right to assess and score only the top-ranked N non-NIL responses in each submission file, where N is determined by assessing resources and the total number of responses returned by all participants.

3.3 Particular Cases

3.3.1 per:alternate_names

The per:alternate_name slot needs separate treatment because systems may extract it without any contextual information (other than occurrence in the same document). While textual patterns may sometimes provide useful context for this slot (e.g., "Dr. Jekyll, *also known as* Mr. Hyde"), it is possible to extract instances of this slot without such information. For example, a system may decide that "IBM" is an alternate name for "International Business Machines" solely based on the fact that the former is an acronym for the latter and they appear in the same document. To allow for these situations, we will accept empty justifications for this slot. In other words, Column 8 may contain an empty string if no textual context was used for the extraction of a per:alternate_names instance. However, the rest of the columns in the output must be populated similarly to the other slot types.

3.3.2 per:title

The definition of the per:title slot changed considerably this year. The main difference is that titles that represent positions at different organizations must be reported as distinct fillers. 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)

These positions must be reported as distinct titles by the systems. Note that this is different from the past evaluations. In the previous evaluations, these titles would be merged into a single instance, because the strings ("CEO") are similar. This year, we are considering these as three distinct, valid fillers since they each refer to a different position at different organizations.

Note that this change in specification does not apply to occupations that have no clear affiliation (e.g., "actor", "star") or to positions where the affiliation is missing. In such situations, the systems (and the human assessors) should revert to the matching criterion of the previous year, where the context for the title slot filler is ignored. One more complicated scenario involves multiple positions with affiliation present for only a few. For example, "M. Smith" may appear in a document as "professor at NYU", "professor at Berkeley" or simply as "professor". In such situations, the position without an affiliation must be reported as separate filler, distinct from the ones with explicit affiliation. In the above example, an ideal system would extract three "professor" fillers, one for the position at NYU, one for the position at Berkeley, and a final one for the unaffiliated position.

The provenance for per:title does not change. That is, systems should include offsets for the position string alone, e.g., for "professor" rather than "professor at NYU". However, the justification column (Column 8) must contain the corresponding organization, if present, e.g., the sentence containing "professor at NYU" for the previous example.

Please read the slot annotation guidelines for more details and additional changes in annotation guidelines.

3.3.3 per:employee_or_member_of

This slot merges the former per:employee_of and per:member_of slots. Additionally, the annotation guidelines were simplified: this slot now covers organizations that have hired individuals for temporary positions, e.g., as independent contractors, adjunct or visiting professors, or postdocs, which were not considered before.

Please read the slot annotation guidelines for more details and additional changes in annotation guidelines.

3.4 Scoring

We will pool the responses from all the systems and have human assessors judge the responses. To increase the chance of including answers that may be particularly difficult for a computer to find, LDC will prepare a manual key, which will be included in the pooled responses.

The slot filler (Column 5) in each non-Nil response is assessed as Correct, ineXact, Redundant, or Wrong:

1. A response that contains more than two sentences in the justification (Column 8) will be assessed as Wrong.
2. Otherwise, if the text spans defined by the offsets in Columns 6-8 (+/- a few sentences on either side of each span) do not contain sufficient information to justify that the slot filler is correct, then the slot filler will also be assessed as Wrong.
3. Otherwise, if the text spans justify the slot filler but the slot filler in Column 5 either includes only part of the correct answer or includes the correct answer plus extraneous material, the slot filler will be assessed as ineXact. No credit is given for ineXact slot fillers, but the assessor will provide a diagnostic assessment of the correctness of the justification offsets for the response.
4. Otherwise, if the text spans justify the slot filler and the slot filler string in Column 5 is exact, the slot filler will be judged as Correct (if it is not in the reference KB) or Redundant (if it is in the reference KB). The assessor will also provide a diagnostic assessment of the correctness of the justification offsets for the response.

Two types of redundant slot fillers are flagged for list-valued slots. First, two or more system responses for the same query entity and slot may have equivalent slot fillers; in this case, the system is given credit for only one response, and is penalized for all additional equivalent slot fillers. (This is implemented by assigning each correct response to an *equivalence class*, and giving credit for only one member of each class.) Second, a system response will be assessed as Redundant with the reference knowledge base if it is equivalent to a slot filler in the reference knowledge base; in KBP 2013, these Redundant responses are counted as Correct, but NIST will also report an additional score in which such Redundant responses are neither rewarded nor penalized (i.e., they do not contribute to the total counts of Correct, System, and Reference below).

Given these judgments, we can count:

Correct = total number of correct equivalence classes in system responses
System = total number of non-NIL system responses
Reference = number of single-valued slots with a correct non-NIL response +

number of equivalence classes for all list-valued slots
Recall = Correct / Reference
Precision = Correct / System
F = 2*Precision*Recall/ (Precision + Recall)

The F score is the primary metric for system evaluation.

Summary of important changes since 2012:

- There is a provenance field for the query entity;
- The provenance field for both query entity and slot filler must contain at least one mention and at most two;
- The justification field must contain at least one clause/sentence and at most two sentences, possibly discontiguous.
- The per:alternate_names slot may report an empty justification;
- The per:title slot must report different titles for positions that apply to different organizations. The justification for the per:title slot must include the affiliated organization.
- The per:employee_of and per:member_of were merged into a single slot: per_employee_or_member_of.
- Entities in document meta data can be used as input for the slot fillings tasks or fillers to be extracted by systems. For example, systems should consider as slot filler candidates the post authors, which are recorded in the meta data of discussion forum documents.

4 Temporal Slot Filling

KBP2013 will contain a task of temporal slot filling, which will be based on the temporal slot filling pilot at KBP 2011. The goal of this new task is to add limited temporal information to selected slots in the KBP slot-filling output. We will limit temporal information to the following slot types:

per:spouse
per:title
per:employee_or_member_of
per:cities_of_residence
per:statesorprovinces_of_residence
per:countries_of_residence
org:top_employees/members

For the temporal task, the input query will be a binary relation consisting of an entity and one slot filler. This is different from the 2011 pilot, when the queries consisted of entities alone. With the new format, systems can focus on the temporal aspect of the task, ignoring the slot filling extraction component. The input format for a TSF is very close to the *output* format of the regular slot filling task:

Column 1: query id
Column 2: slot name
Column 3: query entity name (subject of relation)
Column 4: a single docid that justifies the relation between the query entity and the slot filler
Column 5: a slot filler (possibly normalized, e.g., for dates)
Column 6: start-end offsets for representative mentions used to extract/normalize filler
Column 7: start-end offsets for representative mentions used to extract/normalize query entity

Column 8: start-end offsets of clause(s)/sentence(s) in justification
Column 9: confidence score (set to 1.0 for this task)
Column 10: entitykbid
Column 11: fillerkbid

In the above list, we show in bold font the columns that are different from the output format of the regular SF task. Column 1 contains a unique query ID. Note that this query ID is generated specifically for the TSF task and will be the ID for the **relation**, not necessarily related to any regular SF query (which is an **entity**). Column 3 contains the name of the entity, i.e., the subject of the relation described in this query. Columns 10 and 11 contain the IDs in the KBP knowledge base of the entity and filler, respectively. The rest of the columns are identical to the output of the regular slot filling task, with the observation that Column 4 contains a valid document ID (cannot be NIL) and Column 9 contains a confidence score set to 1.0 (meaning that the relation is correct).

For example, a validly formatted query (ignoring the offsets in Columns 6 through 8) for the relation per:spouse(Barack Obama, Michele Obama) is:

Column 1: TEMP70711
Column 2: per:spouse
Column 3: Barack Obama
Column 4: AFP_ENG_20081208.0592.LDC2009T13
Column 5: Michelle Obama
Column 6: XXX-YYY
Column 7: ZZZ-WWW
Column 8: SSS-TTT
Column 9: 1.0
Column 10: E0566375
Column 11: E0082980

For the above query, systems will have to extract the temporal validity of the per:spouse relation between the entity Barack Obama and the slot filler Michelle Obama.

The output for the full temporal task will be scored through system output pooling, like the regular slot filling task.

4.1 Representation of temporal information

Associated with each binary relation will be a 4-tuple of dates

[T1 T2 T3 T4]

indicating that the relation is true for a period beginning at some time between T1 and T2 and ending at some time between T3 and T4. A hyphen in one of the positions implies a lack of a constraint. Thus [- 20110101 20110101 -] implies that the relation was true starting on or before January 1, 2011 and ending on or after January 1, 2011; i.e., that it was true on January 1, 2011 and no further information is available from the texts. Similarly, [20100101 20101231 - -] implies that the relation was true starting at some time in 2010.

The goal in selecting this representation was to be able to capture most of the temporal information conveyed in the text while still retaining the structured data base style of KBP slot

filling. A pair of dates would not be sufficiently flexible – the texts often do not give specific start and end dates. On the other hand, a more general representation involving multiple temporal predicates would be a sharp departure from infobox style.

Some types of information expressed in the text cannot be captured by a 4-tuple. These include:

- Durations where neither endpoint is known ("he worked for IBM for 7 years")
- Relations between slots ("she married Fred two years after moving to Seattle", where the date of the moving event is not specified in the document)
- Relations which are true over multiple disjoint intervals ("Cleveland was President from 1885 to 1889 and from 1893 to 1897")
- Regularly recurring events ("each Friday")
- Fuzzy relations ("lately", "recently") that are encoded with the SET type in TimeML.

Here are some examples of 4-tuple representations, assuming the publication date of the text is January 1, 2001:

Document text	**T1**	**T2**	**T3**	**T4**
Chairman Smith	-	20010101	20010101	-
Smith, who has been chairman for two years	-	19990101	20010101	-
Smith, who was named chairman two years ago	19990101	19990101	19990101	-
Smith, who resigned last October	-	20001001	20001001	20001031
Smith served as chairman for 7 years before leaving in 1991	19840101	19841231	19910101	19911231
Smith was named chairman in 1980	19800101	19801231	19800101	-

Table 3. 4-tuple Representation Examples

Note that these values assume that durations are interpreted as being literally exact. For example, "two years ago" is interpreted as exactly two years ago, not (for example) as between 1½ and 2½ years ago. Though this is unrealistic, it simplified the task and the evaluation. In the case of a slot holding over multiple disjoint intervals, the best response will capture the period from the start of the first interval to the end of the last interval.

4.2 Output format

As for the regular slot-filling task, system output files should be in UTF-8 and contain at least one response for each query (relation). Lines should have the following tab-separated values:

Column 1: query id
Column 2: the slot name
Column 3: a unique run id for the submission

Column 4: one of the strings 'NIL', 'T1', 'T2', 'T3', or 'T4'
Column 5: a document ID
Column 6: start-end offsets for representative mentions used to match/normalize filler
Column 7: start-end offsets for representative mentions used to match/normalize entity
Column 8: start-end offsets of clause(s)/sentence(s) in justification for the relation between entity and filler
Column 9: a normalized date. Note that this date might include "X"s if the information is not fully specified in the document, and must include two hyphens to separate month from year and day from month.
Column 10: provenance of the temporal information

If there is no temporal information for a particular query, a single response line should be generated for the query, with Column 4 containing NIL and the remaining columns empty.

If the query relation has some (1 to 4) temporal constraints, up to 4 response lines should be generated for the query, with Column 4 containing the type of constraint (T1, T2, T3, or T4); Column 5 containing the docid of a document supporting the constraint; Column 6 containing the provenance of the slot filler (similar to the regular SF task); Column 7 containing the provenance of the entity (similar to the regular SF task); Column 8 containing the justification for the relation (similar to the regular SF task); Column 9 containing a normalized date; and Column 10 containing the provenance of the temporal information. Similar to the regular slot filling task, Column 10 should include the offsets for at least one mention, and up to two mentions used for the extraction and normalization of temporal information. For example, if a system extracts the relative date "Wednesday" and normalizes it to "2008-12-31" using the document date from the document below:

<DOC>
<DOCID> AFP_ENG_20081231.0121.LDC2009T13 </DOCID>
<DOCTYPE SOURCE="newswire"> NEWS STORY </DOCTYPE>
<DATETIME> 2008-12-31 </DATETIME>
<BODY>
<HEADLINE>
Thousands protest in Brussels against Israeli action in Gaza
</HEADLINE>
<TEXT>
<P>
Thousands took the streets in Brussels on Wednesday calling for
an end to Israeli bombing of the Palestinian Gaza Strip
…
</DOC>

the system should report the offsets for both "Wednesday" and "2008-12-31" (from the <DATETIME> block) in Column 10.

4.3 Scoring

In order for a temporal constraint (T1-T4) to be valid, the document must justify the query relation (same as for regular English slot filling) AND the temporal constraint.

The simplest scoring scheme would mark each valid temporal constraint as correct or incorrect. However, because the time information provided by the texts may be only approximate, such all-or-nothing scoring is likely to lead to problems. Instead we propose to use a score measuring the

similarity of each constraint in the key and system response. Let the date in the key be k_i and the date in the system response be r_i; let $d_i = |k_i - r_i|$, measured in years. Then the score for the set of temporal constraints on a slot is

$$S(slot) = \frac{1}{4} \bullet \sum_{i=1}^{4} \frac{c}{c + d_i}$$

$$c = \begin{cases} c_{overconstraining}, if\ (i \in \{1,3\} \wedge r_i > k_i) \vee (i \in \{2,4\} \wedge r_i < k_i) \\ c_{vagueness}, otherwise \end{cases}$$

where $c_{overconstraining}$ and $c_{vagueness}$ are two constants (tentatively both set to 1 year) such that errors of that amount get 50% credit. This yields a score between 0 and 1.

The absence of a constraint in T1 or T3 is treated as a value of -∞; the absence of a constraint in T2 or T4 is treated as a value of +∞.

Overall system scores are computed the same way as for regular slot filling (see section 3.3) *except* that, in computing the value of *correct*, we take the sum over all correct slot fills of *S(slot)*.

4.4 Training data

To facilitate system development, we will be providing two sets of training. First, we will distribute the training data for the 2011 temporal slot filling pilot. This data consists of annotations corresponding not only to the final 4-tuples for selected queries but also to intermediate local information regarding temporal constraints. Each instance of a slot value in the text will be annotated with temporal information. If the slot value is associated with a temporal expression representing a date or interval, the annotation will specify the temporal expression, the offset of the expression within the document, its normalized form (for example, for specific dates, its 8-digit yyyymmdd value), and the relation between the slot value and the temporal expression.

We will use a set of seven relations developed for temporal annotation by the DARPA Machine Reading program:

Relation	Role of temporal expression	Example
Beginning	the start time for the slot value	Rob joined GE in 1999
Ending	the end time for the slot value	Rob left GE in 1999
Beg_and_end	the slot value is true exactly for the specified time	Rob was named *linguist of the month* for June 1999.
Within	the slot value is true for at least some portion of the specified time	Rob worked for GE in 1999
Throughout	the slot value is true for all of the specified time	Rob commuted to work from his home in Denver for all of 1999
Before_start	a moment before the start time for the slot value	In 1999, before Rob joined GE, …

After_end	a moment after the end time for the slot value	By 1999 Rob had already left GE

Table 4. Temporal Relation Types

In addition, if the slot is currently true (as of the document date) this will be indicated in the intermediate file. Further details regarding these relations can be obtained from the *TSF annotation guidelines* document, which will be distributed along with the training data.

Each of these relations can be translated into a 4-tuple. The basic guidelines for doing so will be distributed as part of the training data. The information in the individual 4-tuples can then be aggregated across sentences and documents, in the simplest cases by taking the maximum of T1 and T3 values and the minimum of T2 and T4 values. Because – as noted above – the 4-tuples do not capture all the temporal information in the text, the procedure based on the aggregation of 4-tuples will not necessarily produce the most accurate corpus-wide 4-tuple.

Second, we will distribute the outputs of the systems that participated in the 2011 pilot, which will soon be re-annotated by LDC for correctness, according to the current specification.

Summary of important changes since 2012:

- The input query consists of a relation, i.e., a pair of entity and slot filler, rather than an individual entity;
- The provenance for temporal slot filling follows the same format as the provenance for the regular slot filling task, with at least one mention and at most two mentions reported. Reporting the offsets for two mentions is mandatory if the system extracted a relative date and used another (e.g., the document date) for normalization.

5 Data

5.1 Knowledge Base and Source Document Collection

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.
Each entity in the KB 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)

The 'raw' slot names and the values in the reference KB are based on an October 2008 Wikipedia snapshot. To facilitate use of the reference KB a mapping from raw Wikipedia infobox slot-names to generic slots is provided in training corpora.

The source document collection for the KBP 2013 English Slot Filling tasks are composed of documents from the following LDC packages:

1. LDC2011T07: A million documents from English Gigaword Fifth Edition
2. LDC2012E23: TAC 2012 KBP Source Corpus Additions Web Documents
3. Approximately half a million discussion forum posts (TBA)

Only a subset of the English Gigaword collection will included in the official KBP 2013 tasks; the document IDs of those Gigaword documents that are part of the KBP 2013 tasks are listed in LDC catalog item LDC2012E22 (TAC 2012 KBP Source Corpus Additions Newswire Doc-ID Lists).
The following Table 3 presents the profile of the collection of source documents for the KBP 2013 slot-filling tasks.

Language	Source	Genre	Size (documents)
English	LDC2011T07	Newswire	1,000,257 (list in LDC2012E22)
	LDC2012E23	Web Text	999,999
	TBA	Discussion Fora	99,063

Table 5. Distribution of Documents in KBP 2013 Source Document Collection. All these source documents are distributed by LDC as a single corpus: "TAC 2013 KBP Source Corpus" with Catalog ID LDC2013E45

Important note: although, as mentioned, the above source documents are retrieved from various other LDC corpora, they will be distributed by LDC to KBP participants as a single corpus, entitled "TAC 2013 KBP Source Corpus", with Catalog ID LDC2013E45. We hope that this simplifies data management for participants.

5.2 Training and Evaluation Corpus

The following Tables summarize the KBP 2013 training and evaluation data that we aim to provide for participants. For all tasks we try to achieve a balance among genres, and between the queries with and without KB entry linkages.

Corpus	Source	Size (entities)	
		Person	Organization
Training	2009 Evaluation	17	31
	2010 Participants	25	25
	2010 Training	25	25
	2010 Training (Surprise SF task)	24	8
	2010 Evaluation	50	50
	2010 Evaluation (Surprise SF task)	30	10
	2011 Evaluation	50	50
	2012 Evaluation	40	40
Evaluation	2013 Evaluation	50	50

Table 6. English Monolingual Slot Filling Data

Corpus	Size (relations)
Training	Approximately 40 queries for each of the 7 slots
Evaluation	Approximately 40 queries for each of the 7 slots

Table 7. Temporal Slot Filling Data. Note that the amount of training data this year seems larger because we now count relations (i.e., pairs of entities and slot filler) rather than entities.

6 External Resource Restrictions and Sharing

6.1 External Resource Restrictions

As in previous KBP evaluations, participants will be asked to make at least one run (the first 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 additional runs which 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 8.

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. Note that manual annotations of this data are allowed for what is considered world-knowledge (e.g., gazetteers, lists of entities) but only automatically-generated annotations are accepted for KBs of relations that can be directly mapped to slots used in this evaluation.
	Preprocess/annotate a large text corpus before the evaluation to check the correctness of facts or aliases. Same as above, only automatically-generated annotations are accepted for KBs of relations that can be directly mapped to slots used in this evaluation.
Not Allowed	Using structured knowledge bases (e.g., Wikipedia infoboxes, DBPedia, and/or Freebase) to directly fill slots or directly validate candidate slot fillers for the evaluation query
	Editing Wikipedia pages for target entities, either during, or after the evaluation

Table 8. Rules of Using External Resources

6.2 Resource Sharing

In order to support groups that 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 2013 participants will have one week after the evaluation queries are released to return their results for each task. Up to five alternative system runs may be submitted by each team for each task. Submitted runs should be ranked according to their expected score (based on development data, for example). Systems should not be modified once queries are downloaded.

Details about submission procedures will be communicated to the track mailing list. The tools to validate formats will be made available at: http://surdeanu.info/kbp2013/software.php.

7.2 Schedule

Please visit the slot-filling website for an approximate schedule for the English Slot Filling tasks at KBP 2013: http://surdeanu.info/kbp2013/

8 Mailing List and Website

The KBP 2013 website is http://www.nist.gov/tac/2013/KBP/. The website dedicated to the English slot filling tasks is http://surdeanu.info/kbp2013/. Please post any questions and comments to the list tac-kbp@nist.gov. Information about subscribing to the list is available at: http://www.nist.gov/tac/2013/KBP/registration.html.