Task Description for
Knowledge-Base Population at TAC 2013

Spanish Slot Filling

Version 1.0 of May 24, 2013

1 Introduction

The goal of the Knowledge Base Population (KBP) track at TAC 2013 is to promote research in and to evaluate the ability of automated systems that discover information about named entities and incorporate this information into a knowledge source. For the evaluation an initial (or reference) knowledge base (KB) is provided along with a collection of source documents from which systems are to discover information to populate the reference KB. Attributes (a.k.a., "slots") derived from Wikipedia infoboxes are used to create the reference knowledge base. This document provides task guidelines for the KBP 2013 cross-lingual Spanish Slot Filling track (Spanish SF), which involves mining information about entities from Spanish and English text, and populating an English Knowledge Base. For the other tracks in KBP 2013, please visit the KBP 2013 web page: http://www.nist.gov/tac/2013/KBP/.

Spanish slot filling extends the English slot filling task to the cross-lingual paradigm. Given a query entity and a large collection of English and Spanish documents, a system should extract slot fillers for the query in order to populate the reference KB. The 41 slots in the Spanish slot filling task will be the same as those in the KBP 2013 English slot filling task and will consist of both single-valued slots (*e.g.,* per:date_of_birth) and list-valued slots (*e.g.,* per:employee_of, per:children). The KBP 2013 Spanish slot filling task definition closely follows the definition of the KBP 2013 English slot filling task (http://surdeanu.info/kbp2013/def.php), except that Spanish slot filling queries may come from either English or Spanish documents, and slot fillers are extracted from both English and Spanish documents.

Compared to the regular English slot filling track, the Spanish slot filling track aims to achieve the following research goals:

- Extend the slot filling task into a multilingual task by requiring systems to perform the slot filling task by leveraging information found in both English and Spanish source documents. In future years, it is expected that additional languages will be added, but the initial effort on multilingual slot filling is limited to English plus Spanish.
- As with the regular English slot filling track, provide provenance and justification texts that are, at the same time, concise and informative (see Section 2.2 of the regular English slot filling track task definition 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.
- The initial version of the Spanish SF track in 2013 focuses on extraction and linking of slot fillers between English and Spanish documents; future versions of the task may require that slot fillers extracted from Spanish documents be translated into English before being added to the English KB.

2 Input Format

The Spanish slot filling query format is the same as for the KBP 2013 English slot filling task. Each query 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 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 slots which should not be filled. An example query is:

```
<query id="SF_SPA_007">
   <name>Anselmo Moreno</name>
   <docid>APW_SPA_20090503.0878</docid>
   <beg>205</beg>
   <end>218</end>
   <enttype>PER</enttype>
   <nodeid>E0021800</nodeid>
   <ignore>per:date_of_birth per:city_of_birth</ignore>
 </query>
```

The slot guidelines indicate the type of filler for each slot and whether the filler 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).

Note that the KBP 2013 Spanish slot filling task does not require that slot fillers be novel with respect to the reference KB. That is, slot fillers found in the document collection that are already in the reference database *must* be reported as well. We hope this simplifies system development, as developers do not have to check for redundancy with the reference KB. However, slots listed in the <ignore> field must be removed from the submitted output (similar to previous years of English slot filling). For example, for the above query, systems should not extract fillers for per:date_of_birth and per:city_of_birth.

3 Output Format

The system responses take the same form as the KBP 2013 English slot filling task, with slot fillers in the language of the source document that supports the filler. No translation of slot fillers is expected, except for normalization of dates. Along with each slot filler, the system must provide a confidence score, provenance for both filler and query entity, and justification for the relation.

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 information for a slot is believed to be learnable from the document collection, 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 must 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";
- "el 4 de Mayo" would likewise 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 assessment guidelines document (version V3.2 or later) for more details on the normalization requirements.

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 relation extraction. Guidelines for provenance and document offsets are the same as for English slot filling and can be found in the KBP 2013 English Slot Filling Task definition (http://surdeanu.info/kbp2013/def.php).

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: Column 9 must contain a confidence score for the slot filler, following the same guidelines as for the KBP 2013 English slot filling task. *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.*

4 Scoring

Assessment and scoring for the Spanish slot filling task will be the same as for the KBP 2013 English slot filling task. As in the English task, systems should not return more than one response for equivalent slot fillers; two slot fillers are considered equivalent if they refer to the same entity

(in the case of named entities), have the same value (in the case of dates and quantities), or have the same English translation (in the case of strings, such as for per:title and per:charges). The correctness of system responses and the equivalence of slot fillers will be judged by bi-lingual annotators. As for the KBP 2013 English slot filling task, slot fillers that are already in the reference KB will be counted as Correct, but NIST will also report an additional score in which such Redundant responses are neither rewarded nor penalized.

5 Data

5.1 Knowledge Base and Source Document Collection

The reference knowledge base is the TAC KBP 2009 Evaluation Reference Knowledge Base (LDC2009E58[A-C]) and includes 818,741 entities (nodes) based on articles from an October 2008 dump of English Wikipedia. Each entity in the KB includes 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 the October 2008 Wikipedia snapshot. To facilitate use of the reference KB, a partial mapping from raw Wikipedia infobox slot-names to generic slots is provided in training corpora.

The source document collection for the KBP 2013 Spanish Slot Filling task consists of all the English and Spanish documents in the TAC 2013 KBP Source Corpus (LDC2013E45). Table 1 presents the profile of the source documents for the KBP 2013 Spanish slot-filling task.

Language	Genre	Size (documents)
English	Newswire	1,000,257
	Web Text	999,999
	Discussion Fora	99,063
Spanish	Newswire	910,734

Table 1. Documents in TAC 2013 KBP Source Corpus (LDC2013E45) for Spanish slot filling

N.B.: The newswire and web documents in the TAC 2013 KBP Source Corpus are the same as those used in the KBP 2012 entity linking and slot filling tasks, except that older English documents have been excluded (namely, all 1.8 million documents in LDC2010E12: TAC 2010 KBP Source Data); the remaining newswire and web documents come from the following LDC packages:

1. English Gigaword Fifth Edition (LDC2011T07): select document IDs that are part of the KBP 2012/2013 tasks are listed in LDC catalog item LDC2012E22 (TAC 2012 KBP Source Corpus Additions Newswire Doc-ID Lists).
2. TAC 2012 KBP Source Corpus Additions Web Documents (LDC2012E23)
3. Spanish Gigaword Third Edition (LDC2011T12): select document IDs that are part of the KBP 2012/2013 tasks are listed in LDC catalog item LDC2012E22 (TAC 2012 KBP Source Corpus Additions Newswire Doc-ID Lists).

5.2 Training and Evaluation Corpus

The following table summarizes the KBP 2013 training and evaluation data that will be provided to participants.

Corpus	Genre/Source	Size (query entities)	
		Person	Organization
Training	2012 English News, Web; 2012 Spanish News	25	25
Evaluation	2013 English News, Web, Discussion Forum; 2012/2013 Spanish News	40	40

Table 2. Cross-lingual Spanish Slot Filling Data

Note that the Spanish slot filling training data reflects the slot guidelines from KBP 2012, which are slightly different from the guidelines that will be used in 2013 (see Section 2.3 of the KBP 2013 English slot filling definition for a description of the changes since 2012); some slot fillers in the training data may also have been drawn from English documents that were in the KBP 2012 document collection, but that have been excluded from KBP 2013. Additional data from past KBP English slot filling evaluations are also available as training data; see the KBP 2013 English Slot Filling Task definition (http://surdeanu.info/kbp2013/def.php).

Summary of important changes since 2012:

- The requirements for specifying provenance have changed:
  - 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 slots were merged into a single slot: per_employee_or_member_of.
- Addition of English discussion forum posts. Entities in document meta data can be used as input for the slot filling task 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.

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 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 that access the Web. This will provide a better understanding of the impact of external resources.

Further rules for both 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 3. Rules of Using External Resources

6.2 Resource Sharing

In order to support groups that intend to focus on part of the task, 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

Participants will have one week after the evaluation queries are released to return their results. Up to five alternative system runs may be submitted by each team A team may submit some runs that process only English documents or only Spanish documents for the purposes of ablation studies or to compare performance of English vs. Spanish extraction, but at least one run must return both English and Spanish documents. 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://www.nist.gov/tac/2013/KBP/SpanishSF/tools.html

7.2 Schedule

Please visit the KBP 2013 Spanish slot filling website for the schedule for the Spanish slot filling task: http://www.nist.gov/tac/2013/KBP/SpanishSF/

8 Mailing List and Website

The KBP 2013 website is http://www.nist.gov/tac/2013/KBP/. The website dedicated to the Spanish slot filling task is http://www.nist.gov/tac/2013/KBP/SpanishSF/**.** 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.