TAC KBP Slot Filling Assessment

Version 3.1 – July 19th, 2012

Linguistic Data Consortium

Created by: Joe Ellis (joellis@ldc.upenn.edu)

With contributions from: Heather Simpson, Kira Griffitt, Hoa Trang Dang, Ralph Grishman, & Heng Ji

http://projects.ldc.upenn.edu/kbp/

Changes from Last Version

- Additional examples of acceptable fillers for name slots were added under section 2.1

- A description of how to deal with coreference of fillers for PER: Titles and for entities that are "countries of countries" (e.g. the United Kingdom and the European Union).

- Section 2.5 – Justification Assessment Categories – was added.

- Section 2.1.1 – Name Slot Filler Normalizations – was edited to note that normalizations are now required for Name Slot fillers and will have to be assessed.

- Sections 2.1.2 – Value Slot Filler normalizations – was edited to indicate the new format for date normalizations.

Table of Contents

1   Introduction

Text Analysis Conference (TAC) is a series of workshops organized by the National Institute of Standards and Technology (NIST). TAC was developed to encourage research in natural language processing (NLP) and related applications by providing a large test collection, common evaluation procedures, and a forum for researchers to share their results. The Knowledge Base Population (KBP) track of TAC aims to develop systems that can determine whether or not entities have an existing Wikipedia page, extract information about those entities from web and newswire texts, and use the extracted information to populate an existing knowledge base.

In the Slot Filling task, developing systems search a corpus for information about various entities and add any new information to respective infoboxes from a 2008 snapshot of Wikipedia. A Wikipedia **infobox** is a table added to the top right-hand corner of articles that presents a summary of selected facts about the subject of each article. In Wikipedia, the categories of selected facts, or **slots**, are numerous and vary widely depending on the subject of an article. In TAC KBP, however, there are only 42 slots that pertain to general information about persons and organizations.

There are two parts to the Slot Filling Assessment task. Primarily, you will be judging the validity of the responses (**fillers**) and the justifications for them provided during the Slot Filling task. Secondly, you will group together all of the correct, co-referring fillers into **equivalence classes** in order to arrive at a final number of correct and unique responses for each slot per entity, an essential component for scoring system output.

This document will guide you in the assessment of fillers and justifications and in the creation of equivalence classes. Note, however, that in order to correctly complete this task, you will also need a copy of *TAC KBP Slots*, the document which details the subset of Wikipedia infobox slots that systems attempted to fill. Before beginning Slot Filling Assessment, you must familiarize yourself with all of the 42 possible slots (26 for person (PER) entities and 16 for organization (ORG) entities) as they are described in *TAC KBP Slots*.

2   Stage 1 – Assessing Slot Fillers

Sections 2.1 – 2.6 of this document provide detailed guidance on assessing system responses using various pieces of information. However, while you are judging system responses, keep in mind that a filler is generally correct if it is supported by the document from which it was extracted and it meets the requirements for its respective slot as described in *TAC KBP Slots*. While you are assessing system responses, it is also helpful to remember that the basic system task is to add information to a Wikipedia infobox. Keeping the basic task in mind is helpful because, if it is ever unclear whether a filler meets the description of its respective slot, you can ask yourself whether it would be appropriate for inclusion in a Wikipedia infobox.

2.1   Slot Content

Each of the TAC KBP slots are classified as name slots, value slots, or string slots based on the content of their fillers. In addition to classifying the slots, however, the content distinction also serves to guide the assessment of fillers, as detailed below.

Name slots

Fillers for name slots are required to be names, usually that of a person, organization, or geopolitical entity. Although adjectival forms of names are acceptable (e.g., "American", "Christian"), you should reject any fillers that are clearly not names, for instance:

```
PER: Children = five          WRONG – not a name
PER: Spouse = his wife        WRONG – not a name
```

In some cases, systems may return extraneous or incomplete text strings as part of the names that constitute a filler. Following standard practice for Wikipedia infoboxes, fillers that include articles (the, a, & an), titles (Dr., Ms., etc.), or nominal modifiers are acceptable, though not preferred. However, as in the last example below, systems will sometimes include too much extraneous text for a filler, in which case the response should be marked 'wrong' or 'inexact' (see Section 2.4 for details on selecting the appropriate filler assessment category).

Text Excerpt	Acceptable Filler?
the Department of State	acceptable
Department of State	preferred
Rev. Al Sharpton	acceptable
Al Sharpton	preferred
Republican	acceptable
Republican Party	preferred
coach Joe Gibbs	acceptable
Joe Gibbs	preferred
city of Baltimore	acceptable
Baltimore	preferred
the singer-songwriter Hank Williams	acceptable
Hank Williams	preferred
the singer-songwriter Hank Williams who had a string of top hits	unacceptable

2.1.1 Name Slot Filler Normalizations

Developers were required to produce systems that interpret and edit text strings to the most appropriate forms for Wikipedia pages, so occasionally fillers will include normalizations of the text that must be assessed as 'Correct' or 'Wrong'. For example, if your assigned entity was "John Doe" and you found a document containing the text "John Doe's first wife, Ruth", then "Ruth Doe" could be assessed as a correct filler normalization for *PER: Spouse*. Although it is possible that Ruth might not have taken the last name of her husband at the time of their marriage, it is reasonable to assume that she did as long as there is no other information in the document indicating that this may not be the case.

Edited filler text is required if an answer found in a document is correct but the form of the word is unnatural sounding as a knowledge base answer, as is often the case with the adjectival forms of GPE names (e.g., American, Texan, British, etc.). In the table of examples below, note how the edited fillers for *PER: Country of Birth* and *ORG: State or Province of*

Headquarters would be correct while only the unedited text for *PER: Origin* would be. This is because the adjectival form of the word is more appropriate for the *PER: Origin* slot.

Also in the examples below, note that it would not be correct for "The Big Apple" to be edited to the correct name of the GPE it refers to, "New York City". This is because, absent any additional information in the source document, making the connection between the two names for the city would be unsupported:

Slot	Document Context	Correct Answer from Document	Correct Edited Text
PER: City of Death	He passed away last year in his favorite city, the Big Apple	The Big Apple	n/a
PER: Origin	He is American-born	American	n/a
PER: Country of Birth	He is American-born	American	America
ORG: State or Province of Headquarters	The Texan band	Texan	Texas

Value Slots

Value slots are required to be filled by either a number or a date. Number fillers for these slots can be spelled out ("five thousand") or written as a number ("5000") but you should reject any answers that are not values or that cannot be resolved to a value, for instance:

```
ORG: Founded = before he moved        WRONG - not a value
               to this country
```

Keep in mind that valid date fillers will be provided in many different formats, not all of which look like numbers. For instance, if a document states that the assigned person entity was born on "New Year's Day 1985", the filler "New Year's Day 1985" would be acceptable for the *PER: Date of Birth* slot.

2.1.2 Value Slot Filler Normalizations

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).

Full date normalizations are usually calculated using document text and the date on which the document was published or posted. Publication dates for news articles and some web articles will be contained in the Doc ID; however, some web articles will only have a post date contained in the document within the POSTDATE tags. Consider the following examples, each of which provides two different, correct fillers for *PER: Date of Death* for the entity Wesley Posvar (the normalization on the right would be assessed as 'Correct'). Assume for each of these examples that the Doc ID is NYT_ENG_20010802.0034.LDC2007T07 (the first

string of numbers '20010802' indicates that the document was published on August 2nd, 2001):

```
Wesley W. Posvar, former chancellor of the University of Pittsburgh,
died on July 27.
```

          PER: Date of Death – July 27   Normalization – 2001-07-27

Since the text above states that Posvar "died on July 27", the normalization "2001-07-27" would be assessed as correct. Even though the document does not explicitly state that Posvar died in 2001, the year can be reasonably inferred because the article was published on August 2nd of 2001.

```
Wesley W. Posvar, former chancellor of the University of Pittsburgh,
died on Thursday.
```

          PER: Date of Death – Thursday   Normalization – 2001-08-02

Unless stated otherwise in the article, systems can infer that Posvar's death in the above example took place on the Thursday closest to, but not past, the article's publication date. You can check the resolution by referencing a 2001 calendar (either online or using the *cal* command in a UNIX terminal), which shows that 2001-08-02 was actually a Thursday itself, making it the Thursday closest to, but not past, the article's publication date.

```
Wesley W. Posvar, former chancellor of the University of Pittsburgh,
died last year.
```

          PER: Date of Death – last year   Normalization – 2000-XX-XX

Since the above example states that Posvar died "last year" and the article was published in 2001, it is reasonable for systems to assume that the entity died in the year 2000 as long as no conflicting information is provided in the document. Also, note that, since no information is provided on the day or month of Posvar's death, only the year can be correctly inferred.

```
Wesley W. Posvar, former chancellor of the University of Pittsburgh,
died a few years ago.
```

          PER: Date of Death – n/a   Normalization – n/a

In this last example, the text is not informative enough to determine when Posvar died. As a result, neither "a few years ago" nor any resolution drawn from it would be a correct filler.

String Slots

String slots are basically a "catch all", meaning that their fillers cannot be neatly classified as names or values. The text excerpts (or "strings") that make up these fillers can sometimes be just a name, but are often expected to be more than a name.

2.2   Slot Quantity

Single-value

Systems were only supposed to provide a single-filler for all single-value slots. However, since multiple teams participate in the slot filling task and contradictory information could be present in the corpus, it is possible to have more than one valid filler for any single-value slot. Be that as it may, the possibility of multiple, correct, single-value fillers is less likely for some slots (e.g., *PER: Date of Birth*), and more likely for others (e.g., *PER: Age, PER: Religion* or *ORG: Website*).

List Value

Systems were allowed to provide multiple fillers for list-value slots because, for instance, an assigned person entity might have multiple children or have been employed by multiple organizations. Note that list-value slots do not *require* multiple answers, but multiples are permitted.

2.3   TAC KBP Slots Table

Although these guidelines do not include the slot descriptions, the following table of all 42 slots is provided for reference:

Entity Type	Slot Name	Content	Quantity
PER	Alternate Names	Name	List
PER	Children	Name	List
PER	Cities of Residence	Name	List
PER	City of Birth	Name	Single
PER	City of Death	Name	Single
PER	Countries of Residence	Name	List
PER	Country of Birth	Name	Single
PER	Country of Death	Name	Single
PER	Employee Of	Name	List
PER	Member Of	Name	List
PER	Origin	Name	List
PER	Other Family	Name	List
PER	Parents	Name	List
PER	Schools Attended	Name	List
PER	Siblings	Name	List
PER	Spouse	Name	List
PER	State or Province of Birth	Name	Single
PER	State or Province of Death	Name	Single
PER	States or Provinces of Residence	Name	List
PER	Age	Value	Single
PER	Date of Birth	Value	Single
PER	Date of Death	Value	Single
PER	Cause of Death	String	Single
PER	Charges	String	List

PER	Religion	String	Single
PER	Title	String	List
ORG	Alternate Names	Name	List
ORG	City of Headquarters	Name	Single
ORG	Country of Headquarters	Name	Single
ORG	Founded by	Name	List
ORG	Member Of	Name	List
ORG	Members	Name	List
ORG	Parents	Name	List
ORG	Political/Religious Affiliation	Name	List
ORG	Shareholders	Name	List
ORG	State or Province of Headquarters	Name	Single
ORG	Subsidiaries	Name	List
ORG	Top Members/Employees	Name	List
ORG	Dissolved	Value	Single
ORG	Founded	Value	Single
ORG	Number of Employees/Members	Value	Single
ORG	Website	String	Single

2.4   Filler Assessment Categories

In the first task of slot filling assessment, you will mark each filler provided by systems as being 'Correct', 'Wrong', or 'Inexact'.

Correct Fillers

Fillers must meet two requirements in order to be judged as correct. Primarily, all answers must meet the requirements of their respective slots as described in the document *TAC KBP Slots.* Secondly, all fillers must be supported in the documents from which they were selected. If a filler cannot be justified solely by the document from which it was selected, it should not be labeled as correct, even if you know it to be true because of an outside information source (for more information on the appropriate use of outside information sources in assessment, see section 2.6 – *Using Outside Knowledge Sources*).

Wrong Fillers

There are two ways in which slot fillers can be simply wrong. Primarily, all answers must meet the requirements of their respective slots as described in the document *TAC KBP Slots.* As a result, any fillers that do not meet the requirements of their respective slots are wrong.

Secondly, all fillers must be supported in the documents from which they were selected. If a filler cannot be justified solely by the document from which it was selected, it is wrong, even if you know it to be correct because of an outside information source (for more information on the appropriate use of outside information sources in assessment, see section 2.6 – *Using Outside Knowledge Sources*).

Inexact Fillers

A slot filler should be judged as inexact if it meets both of the standards for correct fillers (i.e., it is supported in its source document and fulfills the requirements of its respective slot) but the string of text selected is incomplete, includes extraneous text, or is not the most informative text string in the document that refers to the filler entity. For example, given the entity 'Michelle Obama' and a source document with the text:

```
Barack Hussein Obama is the U.S. President. He was elected in November
2008. Obama and his wife, Michelle, have two daughters…
```

the correct filler for *PER: Spouse* would be "Barack Hussein Obama". The text excerpt "Obama" would be an inexact filler because "Barack Hussein Obama" would be the most informative answer in the document. Note that source documents from the web will occasionally contain entities' names in strange or informal formats. Such text excerpts are acceptable slot fillers and would only be inexact if another, more correct or informative version of the name appeared elsewhere in the document. Here are some additional examples:

Slot	Document Text	System Answers	Assessment
PER: Siblings	His sister Emily… (no other mention of Emily in document)	Emily	Correct
PER: Siblings	Emily Smith, his adopted sister, was quick to support him. Emily has run a foundation for …	Emily	Inexact ("Emily Smith" would be the correct filler)
ORG: Founder	Microsoft Founder Bill Gates is one of the world's most famous billionaires. Gates started his empire…	Microsoft Founder Bill Gates	Inexact ("Bill Gates" would be the correct filler)
PER: Member of	Blanton was a first-round pick for the A's in the 2002 entry draft. Oakland received the pick from the New York Yankees as compensation for the loss of Jason Giambi	Oakland	Inexact ("A's" would be the correct filler)

2.5   Justification Assessment Categories

Justification for a filler is the full string of text that proves the relationship between an entity and a filler. These strings of text can contain just a few words but can also include a few sentences. In either case, the related filler and some mention of the entity should be included in the justification. Just like slot fillers, the three assessment categories for justification are Correct, Wrong, and Inexact.

Note that justification is provided by systems only to assist in the assessment process. If the justification for a filler is wrong or inexact, you must still review the filler's source document in order to determine its validity separately.

Correct Justifications

In order for a filler justification to be correct, it must contain all the information required to link the entity to the filler by the chosen slot.

Wrong Justifications

If a justification is wrong, it does not include any pieces of information necessary to link the entity to the filler by the chosen slot

Inexact Justifications

Strings of text that contain part, but not all, of the information necessary to link an entity to the relative filler by the chosen slot should be assessed as Inexact.

2.6   Using Outside Knowledge Sources

Occasionally, you will learn or already know about a slot filler for the targeted entity through online searching, your own knowledge, and/or your entity's reference document. Note that you **cannot** use this information to assess a filler without first checking whether the answer is supported in the source document. Remember, slot fillers are only correct if they can be justified solely by the source documents from which they were extracted.

For example, you might already be aware that Michelle Obama is President Obama's wife and so, if "Barack Obama" were your assigned entity, and "Michelle Obama" was listed as a filler for *PER: Spouse*, you might be tempted to just mark it as correct. However, if the source document from which the text string "Michelle Obama" was extracted did not include any lexical clues to indicate that she was a spouse of the targeted entity (e.g., "marriage", "wife", "first lady", etc.), the filler **must** be marked as 'Wrong'.

Another common temptation for using world knowledge involves extension of geo-political entities for residence, birth, death, or headquarters slots. For example, if you had already labeled "Texas" as a correct filler for *PER: State or Province of Birth* for your assigned entity, "The United States" could not automatically be labeled as a correct filler for *PER: Country of Birth* unless its source document clearly indicated that Texas was located within the United States or, more simply, that the entity resided in the United States.

Conversely, if you have to make a judgment on a filler that you know to be incorrect, **do not** mark it wrong without first checking to see if it is justified in its source document. For example, if "George Bush" was returned as a filler for *PER: Spouse* for Barack Obama and the source document stated that "Barack Obama married George W. Bush in 2008" without any additional information in the document to indicate that the statement was figurative or untrue, the filler should be marked correct.

Although you cannot generally rely on outside knowledge to assess slot fillers, you can use it to clarify whether appropriate slots were selected for supported fillers. For example, imagine that the targeted entity "Abdurrahman Wahid" had "South Jakarta" provided as a filler for *PER: Cities of Residence* and the source document stated that "Abdurrahman returned to his house in Cilandak, South Jakarta, Indonesia". While this sentence clearly supports one or more residence relations between Wahid and the named locations, you might not know the geo-political level of Cilandak or South Jakarta. In such a case, it would be acceptable to perform an online search, which would reveal that South Jakarta is a city (and that the filler was correct) and that Cilandak is a subdistrict of the city (and, thereby, wrong for any city slot).

3   Stage 2 – Creating Equivalence Classes

Throughout the corpus, all entities mentioned could be referred to by many different names (e.g. "Hillary Rodham Clinton" might be referred to as "Hillary", "Hillary Clinton", "Senator Clinton", "Secretary of State Clinton", etc.). As any of these names could have been marked correct during the first phase of assessment, your job in the second stage is to identify these coreferential fillers for each slot and cluster them together into equivalence classes (in the preceding example, all the different names for "Hillary Rodham Clinton" would be grouped together into a single entity equivalence class). This step is necessary because it provides a total number of correct **and** unique answers per slot for each entity and because it indicates whether systems returned any redundant fillers. Note that, in order for two fillers to be considered coreferential, they must refer to the same entity; they **cannot** be simply related. Consider the following *ORG: Country of Residence* examples for the entity FirstGroup:

```
"Britain's biggest bus firm FirstGroup..."

"UK's FirstGroup agrees to buy Laidlaw"
```

Given the text, both "Britain" and "UK" would be valid fillers for *ORG: Country of Residence*. Since the UK operates as a "country of countries", which includes Britain, it is likely that both of these fillers are referring to the same location of FirstGroup's headquarters. However, since the UK and Britain are not strictly the same entity, the two fillers should occupy separate equivalence classes. Note though that, following the guidelines for slots such as *PER: Origin* and *PER: Country of Birth*, there could be cases in which you would place adjectives and nouns into the same equivalence classes (e.g. "French" and "France").

Because any slot could potentially have more than one correct answer (particularly list-value slots), you will likely have to create multiple equivalence classes, one for each unique entity. For instance, the entity "Michael Jackson" might contain the following correct fillers for the *PER: Children* slot:

Prince Jr.	"Blanket" Jackson
Prince Michael Jackson, Jr.	Paris
Prince Michael "Blanket" Jackson II	Prince Michael Jackson II
Paris Katherine Jackson	Blanket
Paris Jackson	Prince

Each of the above names would be connected to a source document and so, after reading each in context and determining who was being referred to, you would be able to create three equivalence classes, one for each distinct entity mentioned:

Class 1	Class 2	Class 3
Prince Jr.	Paris Katherine Jackson	Prince Michael "Blanket" Jackson II
Prince Michael Jackson, Jr.	Paris Jackson	"Blanket" Jackson
Prince	Paris	Prince Michael Jackson II
		Blanket

While you should primarily rely on information contained in the documents when creating equivalence classes, you may utilize outside information sources to help make your determinations. For instance, if you found that "Blanket" was a nickname for "Prince Michael Jackson II", then you could cluster "Blanket" and "Prince Michael Jackson II" into the same equivalence class, even if the given source documents did not state the information explicitly. Note, however, that if the information contained in the source documents contradicts outside knowledge, you should cluster fillers based on information in the source documents.

3.1   Equivalence Classes for Titles

Fillers for *PER: Titles* present a number of challenges to the decision-making processes involved in creating equivalence classes. As a result, you should keep the following items in mind while coreferencing these fillers:

- Exact or nearly-exact string matches should be coreferenced (e.g. "chief executive" & "chief executive officer" should be considered equivalent)

- Acronyms or common abbreviations should be coreferenced with fuller versions (e.g. "CEO" & "chief executive officer" should be considered equivalent)

- Common word re-orderings should be placed in the same equivalence classes (e.g. "Finance Minister" and "Minister of Finance" should be coreferenced)

- Nearly synonymous terms should be placed in the same equivalence classes (e.g. "attorney" and "lawyer" would be coreferenced as would "Premier" and "Prime Minister")

- Specified and unspecified positions should NOT be coreferenced (e.g., "prosecutor", "attorney", and "U.S. Attorney" would all go into separate equivalence classes).

4   Steps for Slot Filling Assessment

Sign in to the online assessment tool using the URL and email/password combination provided by your supervisor. Afterward, continue to repeat the following steps for each set of slot fillers for your assigned entity until you get a message stating that you have completed all assessment tasks for your assigned entity:

1. **Review entity reference document**
   When you first open the slot filling tool, the name of your assigned person or organization will appear in the upper-left corner of the tool and a reference document for the entity will be displayed in the top right panel. You should review the reference document to get a sense of who or what your entity is before attempting to assess slot fillers for it.

2. **Review slot description**
   Assessment files are formatted so that you will review all of the fillers provided for a particular slot in one batch. Whenever you are given a set of fillers to assess, the relative slot will be listed in the upper-left corner of the tool, next to the name of your assigned entity. Before attempting to assess the fillers, you should take a moment to review the relative slot description in *TAC KBP Slots* (see your supervisor if you do not have a copy of this document for reference during the slot filling assessment task).

3. Assess each slot filler and justification

The panel on the left side of the assessment tool contains all of the fillers provided by systems for your particular slot/entity assignment. When you click on one of the fillers, the source document from which it was extracted appears in the lower right panel of the tool. Read through the source document in search of support for the filler, keeping in mind the description of the respective slot from *TAC KBP Slots*. Based on what you are able to find in the source document, determine whether the filler is 'Wrong', 'Inexact', or 'Correct' and select the appropriate radio button below the filler. Repeat this step until you have recorded a judgment for all of the fillers displayed.

> **HINT:** At the top of the list of fillers are six radio buttons that allow you to filter the list based on judgments you have already made. Before proceeding to Phase 2 – Creating Equivalence Classes – you should select the 'Not Judged' radio button to ensure that you have not missed any fillers.

After you have recorded a judgment for each filler in the left panel, the 'Move to Phase 2' button in the upper-right corner of the tool will become active. Pressing the button will reconfigure the tool to enable equivalence class creation.

4. Create equivalence classes

Click on the first filler in the panel on the left side of the tool and read the relevant sections of the source document to determine who or what the filler refers to. Afterward, proceed with one of the two following steps depending on whether the referent requires a new equivalence class or needs to be added to an existing class (note that there will be no existing entity classes when starting the second phase of assessment and so your first entity will require the creation of a new class):

a) **For entities that require a new class:** If you wish to create your own label for the entity class being created, press 'New Class' while the filler is highlighted and type your chosen name in the space provided. If you wish for the new entity class to simply have the same name as the filler, press 'Create from Filler' instead. Both options will create a new entity class with a unique ID number for adding other fillers to the class if necessary.

b) **For entites being added to an existing class:** While the filler is highlighted, simply press the ID number next to the existing entity class to which you wish to add the filler.