File name: ie-task-def

Information Extraction Task Definition

(Version 2.1, 23 AUG 95)

---

1 Overview of Information Extraction Task

1.1 Scenario-Dependent and Scenario-Independent Subtasks

The overall goal of the Information Extraction (IE) task is to provide an evaluation of IE technology with reduced overhead and reduced non-NLP requirements, as compared to recent MUCs. To enforce the requirements for reduced overhead, the participant preparation for the evaluation will consist of two stages. The first stage will be scenario-independent, and will begin well in advance of the evaluation; the second stage of participant preparation, which is scenario-dependent, will start just one month prior to the evaluation.

As currently envisioned, the task to be performed during test week will consist of two subtasks:

SUBTASK 1. Scenario Template evaluation, the so-called "Mini-MUC," is the traditional template-level subtask, where the participants are evaluated on whether the templates contain exactly the instantiated objects and filled slots as specified in the scenario definition (and reflected in the answer key), with penalties for spurious, missing, and wrong objects and slot fills.

SUBTASK 2. Template Element evaluation, the so-called "predefined objects" evaluation. There are three types of Template Element objects, ORGANIZATION, PERSON, and ARTIFACT. One difficulty with the Scenario Template subtask is that it is subject to the "lynchpin" or "keystone" effect, where a decision whether to instantiate an object carries a high penalty if wrong (points off for each slot fill in that object under the All Objects scoring method). We can reduce the lynchpin effect by having a subtask which does not involve scenario-dependent relevance criteria. Furthermore, this subtask is viewed as an interesting exercise in its own right, as the next step up from the aggregation of the Named Entity task and the Coreference task.

For example, for the Template Element evaluation, an ORGANIZATION object and all possible slots defined for that object type are to be instantiated for each organization mentioned in a given text, even if a given Scenario Template task confines itself to organizations which are airplane manufacturers and requires an organization's type but not its location. For the Scenario Template evaluation, only those Template Element object and slot types that appear in the scenario task definition will be tested. ARTIFACT objects are handled somewhat differently from ORGANIZATION and PERSON objects. For ARTIFACT objects, the Scenario Template will define the particular kind of artifact to be reported and the particular slots to be used from the Template Element BNF for ARTIFACT. The Template Element test for ARTIFACT will be limited to that type of artifact and to the scenario-defined ARTIFACT slots.

1.2 Evaluation Stages

STAGE 1 (with the announcement of the evaluation). The participants are given the definitions for the scenario-independent and scenario-neutral template elements (defined in this document). The definitions in this document do not reflect the requirements of any particular scenario. The participants are also given one or more example IE scenario definitions and data sets, similar in nature (but not in content) to the Scenario Template task(s) to be used for the actual evaluation. During stage 1, it is expected that the participants will develop their systems to perform on the Template Element evaluation subtask (especially ORGANIZATION and PERSON objects) and will design their system to be able to accomodate the template design requirements of Scenario Template task definitions to be released during stage 2 of the evaluation.

STAGE 2 (one month prior to test week). The participants are given one or two scenario definitions. During the course of this one-month period, the participants configure their system to produce the appropriate subset of the Template Elements and to produce the higher-level object(s) as defined in the scenario statement. The entire template for any given task is therefore fairly simple, consisting of one or more Template Element objects, only one scenario-specific (high-level) object, and perhaps a relational object. The number of slots (other than pointer slots) that do not come from the set of Template Elements will be five or less.

1.3 Levels of Template Structure

Four levels of template objects are defined:

LEVEL 1 (Template Element). The objects and slots defined in this document. These are generic Template Elements which may play a role in virtually any task scenario. These template elements are not oriented towards any particular task, but instead attempt to capture the sort of information that may be needed for a wide range of tasks. All of these objects are fairly simple and have no relational information (i.e., no pointers to other objects). For a given IE scenario, only a subset of the predefined Template Element objects will be used; in addition, one or more slots might be ignored from the Template Element objects that are used.

LEVEL 2 (Relational Object -- optional). Objects which define a relation between generic Template Elements and scenario-specific ones. These relations are not included in the Template Element objects, for the purpose of generality and simplicity. For example, a Relational object may consist of a pointer to an ORGANIZATION object (generic), a pointer to a PERSON object (generic), a slot representing the role that the person has in that organization (scenario-specific), and, perhaps, a slot containing temporal information (generic).

LEVEL 3 (Scenario Template Object). For each IE scenario, it is envisioned that there will be exactly one scenario-specific object type. It captures the essential relation or event of interest in the task. This object type will have pointers to the Template Element object types appropriate for the task, as well as pointers to any Relational objects defined for the task. It may also contain slots that are defined as part of the Template Element subtask.

LEVEL 4 (Top-Level Template Object). For each text that is relevant to an IE scenario, there will be exactly one Top-Level Template object. It will identify the text and will contain one or more pointers to Scenario Template objects.

2 Scenario-Dependent Subtask: General Description

An IE scenario task is to identify all information in each input text that is defined to be relevant by the task definition in the scenario Fill Rules document, and to construct a representation of the relevant information in the format specified by the BNF.

For any given IE scenario, the following will be provided:

NARRATIVE. Paragraph that briefly describes the scenario topic and the relevance criteria. The narrative will be used by the evaluation designers in formulating a text retrieval query that will return candidate test set documents from the MUC-6 corpus.

BNF DEFINITION. Will include one or more of the Template Element objects or slots defined in this document, plus any scenario-specific and relational objects needed for the scenario. Primarily defines the syntax of the template.

FILL RULES. Will describe the reporting conditions and the semantics of each object and slot. The Template Element objects will have separate minimum conditions and slot descriptions that are available during the first stage of evaluation; additional reporting conditions may be imposed in the fill rules for a particular scenario (e.g., instead of reporting all organizations, a scenario may only require reporting airplane manufacturing companies).

EXAMPLE BASE. A set of N texts with accompanying filled-out templates (for both subtasks).

3 Scenario-Independent and Scenario-Neutral Aspects of IE Task

3.1 Template Definition

3.1.1 BNF

```
/* Top-Level Object -- applies to Scenario Template subtask only */

<TEMPLATE> :=
      DOC_NR:           "NUMBER"^
      CONTENT:          <scenario-specific-object>*
      COMMENT:          " "-

/* Template Element Objects -- apply to Template Element subtask; apply
selectively to Scenario Template subtask */

<ORGANIZATION> :=
      ORG_NAME:         "NAME"-
      ORG_ALIAS:        "ALIAS"*
      ORG_DESCRIPTOR:   "DESCRIPTOR"-
```

```
      ORG_TYPE:          {GOVERNMENT, COMPANY, OTHER}^
      ORG_LOCALE:        LOCALE-STRING {{LOC_TYPE}} *
      ORG_COUNTRY:       NORMALIZED-COUNTRY-or-REGION
                         | COUNTRY-or-REGION-STRING *
      OBJ_STATUS:        {OPTIONAL}-
      COMMENT:           " "-

<PERSON> :=
      PER_NAME:          "NAME"^
      PER_ALIAS:         "ALIAS"*
      PER_TITLE:         "TITLE"*
      OBJ_STATUS:        {OPTIONAL}-
      COMMENT:           " "-

<ARTIFACT> :=
      ART_ID:            "ID"-
      ART_DESCRIPTOR:    "DESCRIPTOR"-
      ART_TYPE:          {{scenario-specific-set-fill}}
      OBJ_STATUS:        {OPTIONAL}-
      COMMENT:           " "-

LOC_TYPE ::
      {CITY, PROVINCE, COUNTRY, REGION, UNK}

/* Template Element Slots -- Apply to Scenario Template subtask only.
Valence is scenario-dependent. */

LOCALE:            LOCALE-STRING {{LOC_TYPE}}

COUNTRY:           COUNTRY | COUNTRY-STRING

DATE:              {BEFORE, AFTER, ON} DATE-EXP
                   | BETWEEN DATE-EXP DATE-EXP

DATE-EXP ::        ([[01-31]]|{EA, MD, LT, EO, BO})[[01-12]][[00-99]YY]
                   | {EA, MD, LT, EO, BO}
                   {FA, WI, SP, SU, 1Q, 2Q, 3Q, 4Q, 1F, 2F, 3F, 4F, FY}
                   [[00-99]YY]
                   | {EA, MD, LT, EO, BO, FA, WI, SP, SU, 1Q, 2Q, 3Q, 4Q,
                         1F, 2F, 3F, 4F, FY}
                   [[00-99]YY]
                   | [[01-12]][[00-99]YY]
                   | [[00-99]]
                   | DESCRIPTOR
```

3.1.2 Fill Format

3.1.2.1 Slot Types

There are four kinds of slots in the template: set fill, string fill, normalized fill, and index fill (pointer). It should be noted that for purposes of scoring, normalized fills and string fills are equivalent, i.e., the scoring software strips off external double quotes from fills for slots that are defined as taking normalized fills or string fills.

SET FILL. To be filled in by selection from a prespecified list of categories defined in the fill rules for a given slot.

STRING FILL. To be filled in with an exact copy of a text string from the article under analysis. The fill may be enclosed in double quotes, if desired. See the “Tokenization Rules” document for information on what counts as a word token in certain special cases.

NORMALIZED FILL. To be filled with a text string that is converted to a canonical form in accordance with the fill rules for a given slot. The fill may be enclosed in double quotes, if desired.

INDEX FILL (POINTER). To be filled with the index of an object, i.e., a pointer to an object. The fill is to be enclosed in angled brackets.

3.1.2.2 Object Identifiers

All objects are identified by the object name (from the template BNF), the document number (from the DOCNO tag in the text), and a one-up number; a dash is used to separate those three elements. For Wall Street Journal articles, the dash internal to the value of DOCNO must be suppressed; thus, a valid ORGANIZATION object identifier for DOCNO 891026-0100 would be <ORGANIZATION-8910260100-1>.

3.1.2.3 Notation Reserved for Use in Answer Keys

Legitimate ambiguity or vagueness in the text is reflected in the answer key by the presence of alternative acceptable fills. The “/” notation is reserved for this use; such fills are *not* to be generated by the system under evaluation. The notation allows the answer key to present alternate acceptable single fills for a slot, alternate sets of fills for a slot, optional fills (one fill or zero fills), and combinations thereof. An object is treated as optional if all pointers to it are either optional or in a list of alternatives.

Since the Template Element subtask does not include the creation of pointers to the template element objects, the optionality of ORGANIZATION, PERSON, and ARTIFACT objects is indicated via the OBJ_STATUS slot within the optional object itself. The OBJ_STATUS slot is not used for the Scenario Template subtask.

The COMMENT slot may contain notes that the analyst wants to record concerning the answer key. The slot is not scored. (Analysts should avoid entering double quotes within the comment, as they will prevent the template-filling tool, Tabula Rasa, from being able to reload the template file.)

3.2 Fill Rules

The input text contains some SGML tags, including TXT; the IE task is to be performed on the text delimited by the TXT, HL, DATELINE, and DD tags. (Note, however, that the DD tag sometimes doesn’t appear at all, sometimes appears once, and sometimes appears twice.)

Lines within the <TXT> portion of the article that start with the “@” sign signify a table or other special line formatting within the text and should NOT be used for extraction. (However, such lines may also appear within the <HL> portion of the article, and these should be analyzed for extractable material.)

3.2.1 TEMPLATE Object

DEFINITION: Top-level object. Applies to Scenario Template subtask only.

MINIMUM INSTANTIATION CONDITIONS: For every Scenario Template, instantiate one TEMPLATE object.

3.2.1.1 DOC_NR Slot

DEFINITION: Article identifier. To be copied from the DOCNO tagged string in the text. Normalize the string to remove any internal dashes, e.g., 870101-0001 becomes 8701010001. This slot is not scored; it is used only to assist people in associating the template with the original article.

3.2.1.2 CONTENT Slot

DEFINITION: Pointer to object that captures info relevant to a given scenario. It is possible for CONTENT to have multiple values, corresponding to different relevant events described. Relevant events are defined as being different when the value of one slot in the scenario object is incompatible with the value of another.

MINIMUM INSTANTIATION CONDITIONS: Depends on scenario definition.

3.2.2 ORGANIZATION Object

DEFINITION: Corporate, governmental, or other kind of organization.

MINIMUM INSTANTIATION CONDITIONS: Text must refer to a particular organization and must provide fill for at least one of the following slots: ORG_NAME, ORG_DESCRIPTOR.

3.2.2.1 ORG_NAME Slot

DEFINITION: The proper name of the organization, including any corporate designators (see reference document titled "Table of Corporate Designator Abbreviations"). If a document contains more than one variant of the name, the ORG_NAME slot is to be filled with the most complete variant.

MINIMUM INSTANTIATION CONDITIONS: The name must appear in the text.

SPECIAL USAGE NOTES:
1. This slot has a 0 or 1 valence to allow the situation where an unnamed organization participates in an event (or relation) of interest and is perhaps referenced only by a descriptive phrase.
2. If an organization is changing name, report the current name as ORG_NAME and the past or future name as ORG_ALIAS.
3. See "Named Entity Task Definition" for information on treatment of names such as "McDonald's of Japan."

3.2.2.2 ORG_ALIAS Slot

DEFINITION: Variant of the proper name entered in the ORG_NAME slot. There may be more than one value for this slot.

MINIMUM INSTANTIATION CONDITIONS: The variant must appear explicitly in the text. This slot can be filled only if ORG_NAME is filled also.

SPECIAL USAGE NOTES:
1. Misspelled variants of the name reported in ORG_NAME are to be reported in ORG_ALIAS.
2. If the organization is involved in a name change, report the current name as ORG_NAME, and the past or future name as ORG_ALIAS.

3.2.2.3 ORG_DESCRIPTOR Slot

DEFINITION: Noun phrase describing or referring to an organization without naming it. This slot is not permitted to have more than one value.

MINIMUM INSTANTIATION CONDITIONS: Text must provide a string that describes the organization and that does not fit the definition of the ORG_NAME slot. Strings that are used in the article to describe a set of organizations are not candidates for this slot, e.g., "the two new subsidiaries."

SPECIAL USAGE NOTES:
1. This slot is intended to capture information on the organization other than its name or alias. Therefore, the string fill for this slot is not permitted to contain the name or alias, which means that the fill will sometimes be a substring of a full noun phrase. The substring could be a premodifier noun or noun phrase or a head noun or noun phrase; it cannot be a non-NP (e.g., cannot be a possessive, prepositional phrase, or a pure adjective). Below are a few examples of complex NPs with descriptor substrings:

"the law firm Smith Blarney" (descriptor is "the law firm")
"ABC Corp's XYZ subsidiary" (descriptor is "subsidiary"

2. The answer key will not contain any "insubstantial" descriptors, which includes pronouns (e.g., "it") and simple noun phrases whose head is one of the following nouns:

"administration"
"agency"
"board"
"committee"
"company"
"concern"
"corporation"
"firm"
"government"

"institution"
"unit"

By "simple noun phrases," we mean ones that consist only of the bare head and ones that are modified only by a determiner (e.g., "the," "his," "this") or by an optional determiner and a proper noun string containing the name/alias of the company in question. Taking the word "unit" as an example, the following usages of it would be regarded as insubstantial, where the expressions constitute the complete NP:

"unit" (as a bare noun, perhaps in a headline)
"the unit"
"his unit"
"that unit"
"the ABC Corp. unit" (where "unit" refers to "ABC Corp.")
"XYZ Corp.'s ABC Corp. unit" ("unit" refers to "ABC Corp.")

As a consequence of this guideline, an ORGANIZATION object will not be instantiated if the text provides no name and if the only descriptive information on it is an insubstantial descriptor.

3. All other descriptive noun phrases will be included as alternatives in the answer key. Thus, even the "insubstantial" head nouns listed above may occur in substantial noun phrases. For example, the following usages of "unit" would be regarded as substantial, and the entire phrase would be generated as ORG_DESCRIPTOR:

"a unit of ABC Corp."
"the ABC Corp. unit" (where "unit" refers to an org *other* than "ABC Corp.")
"the new unit"
"the New York unit" (even though "New York" would also appear in ORG_LOCALE)
"the unit based in New York"

4. The answer key will contain alternative fills when the full NP contains one of the following types of modifiers/adjuncts, which are considered to be either of no interest to the database or of questionable parse and limited interest to the database:

a. possessive pronoun premodifier, e.g., "its most profitable subsidiary" (alternate fill is "most profitable subsidiary")

b. temporal adverbials, e.g., "now the most profitable subsidiary"(alternate fill is "most profitable subsidiary")

c. loose adjunct, e.g., a nonrestrictive relative clause or similar type of full or reduced clause, as in "the profitable subsidiary, which announced increased earnings again this quarter" (alternative fill is "the profitable subsidiary") or "the profitable subsidiary, being the second-smallest of all the company's subsidiaries" (alternative fill is "the profitable subsidiary")

5. To qualify as a descriptor, the noun phrase does not have to be definite (e.g., it may be modified by the indefinite article "a"). Thus, the phrases enclosed between asterisks in the following examples are allowable fills for ORG_DESCRIPTOR for General Dynamics Corp.:

"*A major government contracting firm* announced today that it has won a new contract. General Dynamics Corp. said..."

"General Dynamics Corp. is *a major government contracting firm*."

"General Dynamics Corp., *a major government contracting firm*, ..."

3.2.2.4 ORG_TYPE Slot

DEFINITION: Categorization of organization as a corporate entity, a government entity, or some other kind of organizational entity.

MINIMUM INSTANTIATION CONDITIONS: The ORG_TYPE fill should be based on evidence from the text or on world knowledge; the slot should never be left blank.

SPECIAL USAGE NOTES:

1. The categories that are to be used for ORG_TYPE are defined as follows:

COMPANY -- any profit-making or nonprofit legal (usually) entity, including universities, partnerships, corporations, proprietorsips, consortiums, enterprises, government-owned corporations, etc.

GOVERNMENT -- the government of a country, state, municipality, etc., or government body such as a government ministry, agency, commission, or committee. In the case of a string such as "IBM announced a joint venture with China," report "China" as type GOVERNMENT unless there is evidence for a different type elsewhere in the text.

OTHER -- organizational entities that do not fit the above categories, such as "the Apache Indian tribe," "OPEC," "the Medellin cartel," "NATO."

3.2.2.5 ORG_LOCALE Slot

DEFINITION: Specific place where an organization is located. Only the most specific place is to be reported. (This will enable accurate, automatic scoring.) The literal string that appears in the text, plus a categorization of the place name, appear in this slot as a complex (two-part) fill. DO NOT ENCLOSE THE LITERAL STRING IN DOUBLE QUOTES.

MINIMUM INSTANTIATION CONDITIONS: The locale name must be specifically mentioned in the text in either noun or adjective form. NOTE: Except in the case of organizations of type GOVERNMENT, the name itself is not to be used as a source of information for the ORG_LOCALE slot.

SPECIAL USAGE NOTES:

1. NAMES

a. The "MUC-6 Reference Gazetteer" does not contain an exhaustive list of the place names that may be used to fill the ORG_LOCALE slot, nor does it usually provide alternative spellings for place names. If the place name is given in the text in adjective form, e.g., "Philadelphian," and does not appear anywhere in the text in noun form, e.g., "Philadelphia," report the name in adjective form.

b. If the text provides only a relative locale such as "near Tokyo" or "60 miles from Tokyo", report "Tokyo" as ORG_LOCALE name.

2. TYPES

a. The location categories that are to be used for ORG_LOCALE are defined as follows:

CITY -- a town, city, port, suburb, or other local settlement

PROVINCE -- a state, province, island or similar subnational geographically or politically defined area

COUNTRY -- a nation, country, colony, federation of countries such as the Confederation of Independent States (the former USSR), or other similar national entity

REGION -- an international region such as Eastern Europe, the Pacific Rim, or the Malay Archipelago

UNK -- a location whose possible type cannot be identified from evidence in the text or from world knowledge. Use UNK as locale type only if the type cannot be determined from the text.

b. The "MUC-6 Reference Gazetteer" uses more location categories than are to be reported in ORG_LOCALE. The following mappings apply:

PORT and AIRPORT in gazetteer are to be reported as CITY in ORG_LOCALE.

ISLAND in gazetteer is to be reported as PROVINCE in ORG_LOCALE.

ISLAND-GROUP in gazetteer is to be reported as either PROVINCE (if part of a single country) or as REGION (if part of an international region).

CONTINENT in gazetteer is to be reported as REGION in ORG_LOCALE.

3.2.2.6 ORG_COUNTRY Slot

DEFINITION: The country or region in which ORG_LOCALE is located. A defining list of names in contained in "MUC-6 Country and Region List." (This list contains only canonical forms. NLP system developers must define their own mappings from the "MUC-6 Reference Gazetteer" and/or other gazetteer resources to this list.)

MINIMUM INSTANTIATION CONDITIONS: To be filled if ORG_LOCALE is filled, even if fill must be inferred. Also to be filled if country can be inferred from certain other text expressions (see item 5 under Special Usage Notes, below).

SPECIAL USAGE NOTES:

1. If ORG_LOCALE is filled in by a name of type COUNTRY or REGION, report the name in this slot as a normalized form drawn from "MUC-6 Country and Region List".

2. Adjective forms such as "Asian" and "Japanese" should be mapped to the noun form on the list, and the noun form should be used as the slot fill.

3. Note that the "MUC-6 Country and Region List" may not contain a complete list of countries and regions. If a canonical form for the name of the country or region does not appear on the list, report the name in noun or adjective form (whichever appears in the text) as a string fill.
4. As a default, assume that "American" refers to "United States."
5. Certain text expressions that indicate an organization's country, such as "the domestic" and "the nation's" in the examples below, occasion the ORG_COUNTRY slot to be filled, if the country referent can be inferred.

"the domestic" <org> /* "the domestic company" */
"the nation's" <org> /* "the nation's largest carrier" */

3.2.3 PERSON Object

DEFINITION: An (unincorporated) person or family.

MINIMUM INSTANTIATION CONDITIONS: Text must supply fill for PER_NAME slot. The guidelines for instantiating a PERSON object are the same as the guidelines given in "Named Entity Task Definition" for annotating person names.

3.2.3.1 PER_NAME Slot

DEFINITION: The proper name of the person or family.

MINIMUM INSTANTIATION CONDITIONS: The text must supply a person or family name.

3.2.3.2 PER_ALIAS Slot

DEFINITION: Variant of the proper name reported in the PER_NAME slot. There may be more than one value for this slot.

MINIMUM INSTANTIATION CONDITIONS: The variant must appear explicitly in the text. This slot can be filled only if PER_NAME is filled also.

SPECIAL USAGE NOTES:
1. Misspelled variants of the name reported in PER_NAME are to be reported in PER_ALIAS.

3.2.3.3 PER_TITLE Slot

DEFINITION: An innate title such as "Dr." or "Ms.," as distinct from a person's role such as "President" or "CEO." (The latter would be captured by a scenario-specific template element such as a Relational object.)

MINIMUM INSTANTIATION CONDITIONS: To be reported only if PER_NAME is filled. The text must explicitly mention the person's title.

3.2.4 ARTIFACT Object

DEFINITION: A product or natural commodity. The nature of the specific artifact(s) to be reported is task-dependent and is therefore defined for a given Scenario Template subtask in the scenario task documentation.

MINIMUM INSTANTIATION CONDITIONS: The text must supply a fill for at least one of the following slots: ART_ID, ART_DESCRIPTOR.

3.2.4.1 ART_ID Slot

DEFINITION: A unique identifier for the artifact.

MINIMUM INSTANTIATION CONDITIONS: Depends on scenario definition.

3.2.4.2 ART_DESCRIPTOR Slot

DEFINITION: Noun phrase describing or referring to an artifact without naming it. This slot is not permitted to have more than one value.

MINIMUM INSTANTIATION CONDITIONS: Text must provide a string that describes the artifact and that does not fit the definition of the ART_ID slot. The string cannot be a pronoun, e.g., "it."

SPECIAL USAGE NOTES:
1. The answer key will provide alternative correct answers if the text supplies more than one substantive descriptor string. If the text provides only uninformative descriptors, e.g., "the product," the fills in the answer

key will all be marked as optional.

3.2.4.3 ART_TYPE Slot

DEFINITION: A categorization of the artifact. Inventory of categories depends on scenario definition.

MINIMUM INSTANTIATION CONDITIONS: Depends on scenario definition.

3.2.5 Template Element Slots

DEFINITION: Task-independent slots (location and time data) that are separate from the predefined objects. They may be defined selectively for a given scenario, e.g., to provide the location and time of an event.

MINIMUM INSTANTIATION CONDITIONS: Depends on scenario definition.

SPECIAL USAGE NOTES:
1. These slots will not be part of the Template Element evaluation. Instead, one or more of them may play a role in one or more Scenario Template subtasks. In such cases, their role will be defined in the scenario task documentation.

3.2.5.1 LOCALE Slot

DEFINITION: Specific locale of an entity or event.

MINIMUM INSTANTIATION CONDITIONS: Depends on scenario definition.

3.2.5.2 COUNTRY Slot

DEFINITION: Country locale of an entity or event.

MINIMUM INSTANTIATION CONDITIONS: Depends on scenario definition.

3.2.5.3 DATE Slot

DEFINITION: An absolute or relative date or date range.

MINIMUM INSTANTIATION CONDITIONS: Depends on scenario definition.

SPECIAL USAGE NOTES:
1. The YY option and DESCRIPTOR option are to be used only if the article contains no DD tags. Use YY if only a partial date is given in the text, e.g., “on 27 March;” the output of extraction for that example would be “ON 2703YY”. Use descriptor string option if a time phrase is used that cannot be represented in the usual date format; for example, “last week” (“ON last week”) or “Tuesday” (“ON Tuesday”).
2. See separate documentation titled “A Revised Template Description for Time (v3)” and “Supplement to Time Treatment Used for MUC-5” for further information.

APPENDIX A. Example of Template Element Objects

These are examples of fills for the Template Element subtask that are extracted from a text in the MUC-5 corpus. The extracted information and the text itself appear below. The ARTIFACT object assumes a scenario-specific IE task that includes sports equipment such as golf clubs (and excludes such things as "golf club parts").

```
<ORGANIZATION-0592-1> :=
      ORG_NAME:          "BRIDGESTONE SPORTS CO."
      ORG_ALIAS:         "BRIDGESTONE SPORTS"
                         "BRIDGESTON SPORTS"
      ORG_DESCRIPTOR:    "JAPANESE SPORTS GOODS MAKER"
      ORG_TYPE:          COMPANY
      ORG_LOCALE:        JAPAN COUNTRY
      ORG_COUNTRY:       JAPAN

<ORGANIZATION-0592-2> :=
      ORG_NAME:          "UNION PRECISION CASTING CO."
      ORG_ALIAS:         "UNION PRECISION CASTING"
      ORG_DESCRIPTOR:    "A LOCAL CONCERN"
                         /"CONCERN"
      ORG_TYPE:          COMPANY
      ORG_LOCALE:        TAIWAN COUNTRY
      ORG_COUNTRY:       TAIWAN

<ORGANIZATION-0592-3> :=
      ORG_NAME:          "TAGA CO."
      ORG_DESCRIPTOR:    "A JAPANESE TRADING HOUSE"
                         /"A COMPANY ACTIVE IN TRADING WITH TAIWAN"
      ORG_TYPE:          COMPANY
      ORG_LOCALE:        JAPAN COUNTRY
      ORG_COUNTRY:       JAPAN

<ORGANIZATION-0592-4> :=
      ORG_NAME:          "BRIDGESTONE SPORTS TAIWAN CO."
      ORG_TYPE:          COMPANY
      ORG_DESCRIPTOR:    "A JOINT VENTURE"
      ORG_LOCALE:        KAOHSIUNG CITY
                         /KAOHSIUNG PROVINCE
      ORG_COUNTRY:       TAIWAN
      COMMENT:           "'A JOINT VENTURE' is the most substantive descriptor"
                         /"In the judgment of the analyst, the locale
'KAOHSIUNG' matches either 'Kao Hsiung' or 'Kao-hsiung' in the 'MUC-6
Reference Gazetteer.'  The former is listed as type PORT and the latter is
listed both as type CITY and type PROVINCE.  Since PORT is collapsed with CITY
as far as the IE task is concerned, that leaves two alternative correct
answers in the answer key."

<ARTIFACT-0592-1> :=
      ART_DESCRIPTOR:    "GOLF CLUBS"
                         /"IRON AND 'METAL WOOD' CLUBS"
      COMMENT:           "ART_TYPE not specifiable without rest of task"
                         /"'UNITS' and 'LUXURY CLUBS' are not viable
alternatives to the fill for ART_DESCRIPTOR (they do not convey as much useful
info)"
```

<doc>
<DOCNO> 0592 </DOCNO>
<DD> NOVEMBER 24, 1989, FRIDAY </DD>
<SO> Copyright (c) 1989 Jiji Press Ltd.; </SO>
<TXT>

BRIDGESTONE SPORTS CO. SAID FRIDAY IT HAS SET UP A JOINT VENTURE IN TAIWAN WITH A LOCAL CONCERN AND A JAPANESE TRADING HOUSE TO PRODUCE GOLF CLUBS TO BE SHIPPED TO JAPAN.

THE JOINT VENTURE, BRIDGESTONE SPORTS TAIWAN CO., CAPITALIZED AT 0 MILLION NEW TAIWAN DOLLARS, WILL START PRODUCTION IN JANUARY 1990 WITH PRODUCTION OF 20,000 IRON AND “METAL WOOD” CLUBS A MONTH. THE MONTHLY OUTPUT WILL BE LATER RAISED TO 50,000 UNITS, BRIDGESTON SPORTS OFFICIALS SAID.

THE NEW COMPANY, BASED IN KAOHSIUNG, SOUTHERN TAIWAN, IS OWNED 75 PCT BY BRIDGESTONE SPORTS, 15 PCT BY UNION PRECISION CASTING CO. OF TAIWAN AND THE REMAINDER BY TAGA CO., A COMPANY ACTIVE IN TRADING WITH TAIWAN, THE OFFICIALS SAID.

BRIDGESTONE SPORTS HAS SO FAR BEEN ENTRUSTING PRODUCTION OF GOLF CLUB PARTS WITH UNION PRECISION CASTING AND OTHER TAIWAN COMPANIES.

WITH THE ESTABLISHMENT OF THE TAIWAN UNIT, THE JAPANESE SPORTS GOODS MAKER PLANS TO INCREASE PRODUCTION OF LUXURY CLUBS IN JAPAN.

</TXT>