MUC-7 Information Extraction Task Definition

Version 4.1

27 October 1997

Nancy Chinchor (chinchor@gso.saic.com)

TABLE OF CONTENTS

1 Brief Definition of Information Extraction Task

Information extraction in the sense of the Message Understanding Conferences has been traditionally defined as the extraction of information from a text in the form of text strings and processed text strings which are placed into slots labeled to indicate the kind of information that can fill them. So, for example, a slot labeled NAME would contain a name string taken directly out of the text or modified in some well-defined way, such as by deleting all but the person's surname. Another example could be a slot called WEAPON which requires as a fill one of a set of designated classes of weapons based on some categorization of the weapons that has meaning in the events of import such as GUN or BOMB in a terrorist event.

The input to information extraction is a set of texts, usually unclassified newswire articles, and the output is a set of filled slots. The set of filled slots may represent an entity with its attributes, a relationship between two or more entities, or an event with various entities playing roles and/or being in certain relationships. Entities with their attributes are extracted in the Template Element task; relationships between two or more entities are extracted in the Template Relation task; and events with various entities playing roles and/or being in certain relationships are extracted in the Scenario Template task.

2 Motivation and Timing

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

2.1 Domain-Independent Information Extraction

No set of texts can literally be said to be domain-independent, but an attempt can be made to define information that represents a generic "domain" or can be said to belong to most domains of interest. The extraction of entities and their attributes as well as facts about those entities constitutes the most domain-independent kind of information extraction. Template Elements such as organizations and persons with some associated attributes can be extracted from texts without knowing what role or relationships those elements may enter into. Template Relations such as organizations and their locations or persons and the organizations they are employed by can be extracted to fill a database that could support decision-making at a later time.

2.2 Domain-Dependent Information Extraction

A set of texts containing a pre-assigned percentage of articles relevant to a domain, i.e.,articles containing well-defined events, can be used for domain-dependent information extraction. The same set of texts can also be processed without regard to the domain for Template Elements and/or Relations. However, the Scenario Template task requires more focused information extraction from only relevant texts about events and only those entities which participate in the events in defined ways.

Scenario Template evaluation is the traditional MUC template-level task, 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 given for spurious, missing, and wrong objects and slot fills. The scenario definition is revealed only one month prior to the day that test responses must be submitted to force limitations on the amount of domain-specific work required of the developers for porting from domain to domain or moving from generic information extraction to domain-specific information extraction.

3 General Description of IE Tasks

3.1 Template Element

Template Element evaluation contains two general Template Element objects: ENTITY and LOCATION. The three types of ENTITY for MUC-7 are ORGANIZATION, PERSON, and ARTIFACT. One difficulty with the Scenario Template task is that it is subject to the "linchpin" 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). We can reduce the linchpin effect by having one or more tasks which do not involve scenario-dependent relevance criteria. Furthermore, this IE task 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 ENTITY object of the ORGANIZATION type 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. 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 in the Template Element task are handled somewhat differently from ORGANIZATION and PERSON objects because there are so many objects mentioned in texts that can be considered artifacts. For ARTIFACT objects, the particular kind of artifact to be reported will be defined and so will the particular slots to be used from the Template Element BNF for ARTIFACT. The Template Element test for ARTIFACT will be limited to that kind of artifact and to that subset of ARTIFACT slots.

For the Template Element task, the following will be provided:

BNF DEFINITION. Will include the Template Element objects and slots defined in the fill rules. Primarily defines the syntax of the objects and slots.

FILL RULES. Will describe the reporting conditions and the semantics of each object and slot. The Template Element objects will have 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 Template Relations (e.g., instead of reporting all organizations, the Template Relation task will only require reporting those which are pointed to by relations) and 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 texts with accompanying filled-out templates.

3.2 Template Relation

Template Relation evaluation contains general relational objects which point to Template Element objects. The three relations included in MUC-7 are LOCATION_OF, EMPLOYEE_OF, and PRODUCT_OF. One limitation of the Scenario Template task coupled with the Template Element task in MUC-6 was that relations between entities had to be encoded as attributes or objects in the template. In the MUC-7 Template Relation task, these relations or "facts" can be broken out on their own level and be separate from the scenario. Furthermore, this task is viewed as an interesting exercise in its own right, as the next step up from the Template Element task and the beginning of a compilation of scenario-independent facts about Template Elements.

For example, for the Template Relation evaluation, an relational object of the LOCATION_OF type and all possible slots defined for that object type are to be instantiated for each organization whose location is specified 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. On the other hand, for the Template Relation evaluation, only those Template Element objects that appear in a Template Relation will be tested. So, if in a given text, an organization is discussed but none of its employees, products, or locations is mentioned, then that organization need not be included among the Template Elements in the Template Relation task. No relational object will point to it, so it can be pruned.

For the Template Relation task, the following will be provided:

BNF DEFINITION. Will include the Template Relation objects and slots defined in the fill rules. Primarily defines the syntax of the objects and slots.

FILL RULES. Will describe the reporting conditions and the semantics of each object and slot. The Template Relation objects will have 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 with their locations, a scenario may only require reporting airplane manufacturing companies)

EXAMPLE BASE. A set of texts with accompanying filled-out templates (for Template Relations and any necessary Template Elements).

3.3 Scenario Template

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-7 corpus.

BNF DEFINITION. Will include one or more of the Template Element and Template Relation objects or slots defined in the companion documents, plus any scenario-specific 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 and Template Relation 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 texts with accompanying filled-out templates (for all three tasks).

4 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 and relations. The definitions of the scenario-neutral template elements and relations 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 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 task (ENTITY objects) and Template Relation evaluation task (relational objects) and will design their system to be able to accommodate 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 scenario definition. During the course of this one-month period, the participants configure their system to produce the appropriate subset of the Template Elements (and perhaps Template Relations, if appropriate) 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 one or more scenario-specific (intermediate-level) objects. Template Relation objects may be used if appropriate. The number of scenario-specific slots (other than pointer slots) that do not come from the set of Template Elements or Relations will be five or less.

STAGE 3 (test week). The participants are given the test texts.

5 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 relation or scenario, only a subset of the predefined Template Element objects will be used.

LEVEL 2 (Template Relation). The relations are objects which define a relation between generic Template Elements. For example, a relational object may consist of a pointer to an ORGANIZATION ENTITY object (generic) and a pointer to a PERSON ENTITY object (generic). The specific role of the employee in the organization will not necessarily be represented at the template relation level. In a scenario template the employee relationship might be indicated and additionally 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 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.

LEVEL 4 (Top-Level Template Object). For each text in the Scenario Template task, there will be exactly one Top-Level Template object. If the text is relevant to an IE scenario, it will identify the text and will contain one or more pointers to Scenario Template objects. If the text is not relevant to the IE scenario, it will identify the text and will contain no pointers to Scenario Template objects.

6 Fill Format

6.1 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.

6.2 Object Identifiers

All objects are identified by the object name (from the template BNF), the document number (from the DOCID tag in the text), and a one-up number; a dash is used to separate those three elements. For all articles, any punctuation or alphabetic characters internal to the value of DOCID must be suppressed; thus, a valid ORGANIZATION object identifier for DOCID nyt960102.0516 would be <ORGANIZATION-9601020516-1>.

6.3 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 task does not include the creation of pointers to the template element objects, the optionality of ENTITY objects is indicated via the OBJ_STATUS slot within the optional object itself. The OBJ_STATUS slot is not used for the Scenario Template task. Template Relations which are optional are also indicated via the OBJ_STATUS slot. If TE’s are optional and referred to by any TR, the scorer automatically make the TR optional. The annotator of the answer key does not need to keep track of the status of lower level objects during the extraction of Template Relations.

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 body of the comment, as they will prevent the template-filling tool, Tabula Rasa, from being able to reload the template file.)

7 Document Sections for Information Extraction

The input texts for MUC-7 from the New York Times News Service contain some SGML tags. Although the articles come from different news publications, the formatting is uniform. The IE task is to be performed on the text delimited by the <SLUG>,<DATE>, <NWORDS>, <PREAMBLE>,<TEXT>, <TRAILER> tags.

8 Description of Task Defining Documents

For MUC-7, there are three companion documents which define the individual tasks: Template Element, Template Relation, and Scenario Template. Each document is updated during the course of the evaluation year and all changes are marked with change bars which refer to the most previous version number of the document. No changes will be made without a change in the version number and date of the document. The IE documents are organized in a FrameMaker book together with this overview, but can be printed separately by those sites who are participating in a subset of the IE tasks.

9 Template Element Task Object Definition

9.1 BNF

```
/* Template Element Objects -- apply to Template Element task; apply
selectively to Template Relation and Scenario Template tasks */

<ENTITY> :=
      ENT_NAME:          "NAME"*
      ENT_TYPE:          {ORGANIZATION, PERSON, ARTIFACT}^
      ENT_DESCRIPTOR:    "DESCRIPTOR"-
      ENT_CATEGORY:      {ORG_GOVT, ORG_CO, ORG_OTHER,
                         PER_MIL, PER_CIV, PER_OTHER,
                         ART_AIR, ART_GROUND, ART_WATER}^
      OBJ_STATUS:        {OPTIONAL}-
      COMMENT:           " "-

<LOCATION> :=
      LOCALE:            "LOCALE"+
      LOCALE_TYPE:       {CITY, PROVINCE, COUNTRY, REGION, WATER, AIRPORT,
                         UNK}^
      COUNTRY:           NORMALIZED-COUNTRY-or-REGION
                         | COUNTRY-or-REGION-STRING -
      OBJ_STATUS:        {OPTIONAL}-
      COMMENT:           " "-

/* Symbols are used as follows: * = 0 or more; - = 0 or 1; ^ = 1;
 + = 1 or more */
```

9.2 Fill Rules

9.2.1 ENTITY Object

DEFINITION: A corporate, governmental, or other kind of organization, an (unincorporated) person or family, or a product which is a vehicle.

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

9.2.1.1 ENT_NAME Slot

DEFINITION: The proper name of the person or family or of the organization, or an identifier for the artifact. Include also variants of the proper name. There may be more than one value for this slot.

MINIMUM INSTANTIATION CONDITIONS: At least one form of the organization/person name or artifact identifier must appear in the text.

SPECIAL USAGE NOTES:
1. This slot has a 0 or more valence to allow the situation where an unnamed entity 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 both the current name as ENT_NAME and the past or future name as ENT_NAME.
3. See “Named Entity Task Definition” for information on treatment of names such as “McDonald’s of Japan.”
4. The guidelines for instantiating a PERSON type are the same as the guidelines given in “Named Entity Task Definition” for annotating person names.
5. Aliases and misspelled variants of the name are also to be reported in ENT_NAME.
6. For organizations, include any corporate designators (see reference document titled “MUC-7 Table of Corporate Designator Abbreviations”).
7. For artifacts include only proper names of vehicles. Series numbers are considered part of the proper name

as are manufacturers (in most cases).

9.2.1.2 ENT_TYPE Slot

DEFINITION: Categorization of entity as an organization, person, or artifact

MINUMUM INSTANTIATION CONDITIONS: The ENT_TYPE fill should never be left blank.

9.2.1.3 ENT_CATEGORY Slot

DEFINITION: Further categorization of entity depending on whether it is an organization, person, or artifact

MINUMUM INSTANTIATION CONDITIONS: The ENT_CATEGORY 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 for an ENTITY of ENT_TYPE ORGANIZATION to be used for ENT_CATEGORY are defined as follows:

ORG_GOVT -- 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 ORG_GOVT unless there is evidence for a different type elsewhere in the text.

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

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

2. The categories for an ENTITY of ENT_TYPE PERSON to be used for ENT_CATEGORY are defined as follows:

PER_MIL -- a person who is a member of the military.

PER_CIV -- an (unincorporated) person who is known or assumed to be a civilian or a family. This fill is the default for persons.

3. The categories for an ENTITY of ENT_TYPE ARTIFACT to be used for ENT_CATEGORY are defined as follows:

ART_AIR -- a vehicle that primarily travels by air or into outer space, such as an airplane, helicopter, or space shuttle.

ART_GROUND -- a vehicle that primarily travels on the ground, such as a car or tank.

ART_WATER -- a vehicle that primarily travels through water, such as a ship or submarine.

Note that artifacts in MUC-7 are limited to vehicles.

9.2.1.4 ENT_DESCRIPTOR Slot

DEFINITION: Noun phrase describing or referring to an entity 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 entity and that does not fit the definition of the ENT_NAME slot. Strings that are used in the article to describe a set of entities are not candidates for this slot, e.g., “the two new subsidiaries,” “both commanders,” or “two civilian 737s.”

Thus, the following types of strings do not qualify as descriptors:

1. Strings that describe a set of persons, organizations, or artifacts, e.g., "Spokesmen for both the union and the company," “ABC Corp. and XYZ Corp. are [partners in a new joint venture].”, where the string in brackets is not a descriptor.

2. Nonspecific references, e.g., the bracketed NP in the following sentences: "That post is typically filled by [a career foreign-service officer who has broad responsibilities for overseeing the daily work of the U.S. diplomatic corps].", “The contract will go to [a minority-owned company].”, and “The Boeing 757 is [a twin-engine, medium- to long-range jetliner that can carry up to 239 passengers, depending on cabin configuration].”

3. References that are only potentially true of the individual, organization, or vehicle, e.g., "X may be an appealing choice for the job," "Y was nominated commissioner," "X was grooming Y to be his replacement," “ABC Corp. may be [a big winner],” “According to early reports, the ill-fated DEF 123 aircraft may have been [a single-engine plane]” and depictions that are denied to be true, e.g., "Z denied that he was a

candidate," "The suspected car was not a foreign made vehicle"
NOTE: There are some borderline cases, where it's not clear whether the predicate ("believed to be", "selection [for]", "emerging[as]" should be understood as factual or speculative/potential:

"[The likely board candidate] is believed to be Mr. Byrne" (Coded as descriptor, but would have been optional if it had been the ONLY descriptor for Mr. Byrne in the text)

"the president's ill-fated selection of Zoe Baird for [attorney general]" (the article later mentions that she has since withdrawn her name -- thus "attorney general" not coded as descriptor)

"X is emerging as [a leading candidate to succeed James Robinson III]" (Coded as descriptor)

"The New York Daily News named Lou Colasuonno to be [editor of the newspaper]." (Coded as descriptor)

SPECIAL USAGE NOTES:
A. For ENT_DESCRIPTOR of organizations:
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"
"squadron"

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 ENTITY 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 ENT_DESCRIPTOR:

"a unit of ABC Corp."
"the ABC Corp. unit" (where "unit" refers to an organization *other* than "ABC Corp.")
"the new unit"
"the New York unit"

"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 ENT_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*, ..."

B. For ENT_DESCRIPTOR of persons:

1. To qualify as a descriptor, a noun phrase must refer either to the person per se or to the person's professional role (or other functional role). Some references to a person's role are made indirectly, e.g., by the use of "as," "work as," "job of":

"The choice of a successor to James Robinson III as [American Express Co.'s chairman]" (bracketed NP is descriptor of "James Robinson III")

"Mr. Tarnoff worked as [a top State Department aide in the Carter administration]."

2. This slot is intended to capture information on the person 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 couple examples of complex NPs with descriptor substrings:

"President Clinton" (descriptor is "President")

"Democratic lawyer Tom Donilon" (descriptor is "Democratic lawyer")

3. The answer key will not contain any "insubstantial" descriptors, which includes personal pronouns and basic titles. As a consequence, a PERSON ENTITY object will not be instantiated if the text provides no name and if the only descriptive information on it is an insubstantial descriptor. The following titles are considered insubstantial: "Mr.", "Ms.", "Miss", "Sir", "Dr.", "Mrs.", "Prof."

4. All other modifying/referring noun phrases will be considered "substantial" and will be included as alternatives in the answer key. In cases of conjoined descriptors, the key will contain alternative fills representing each conjunct separately as well as the conjoined phrase, unless the conjuncts share a modifier.

"Hank Gutman, [[the chief of staff of the congressional Joint Committee on Taxation and [a former Treasury official]]" (three alternative descriptors)

"The choice of a successor to James Robinson III as [American Express Co.'s chairman and chief executive officer]" (one descriptor)

5. 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., "his Army counterpart" (alternate fill is "Army counterpart")

b. temporal adverbials, e.g., "now CEO of United Airlines" (alternate fill is "CEO of United Airlines")

c. loose adjunct, e.g., a nonrestrictive relative clause or similar type of full or reduced clause, as in "a historian at the University of Pittsburgh who helped write the best-selling Pearl Harbor history 'At Dawn We Slept'" (alternative fill is "a historian at the University of Pittsburgh")

6. To qualify as a descriptor, the noun phrase does not have to be marked definite (does not have to be marked with "the"). Thus, the bracketed phrases in the following examples are allowable fills:

"[Another potential choice for a top position in the State Department] is [Democratic lawyer] Tom Donilon"

7. Time-dependent descriptors qualify as fills for the descriptor slot.

"Peter Tarnoff, [president of the Council on Foreign Relations], ... Mr. Tarnoff was [a career foreign service officer from 1961 until 1982]." (two descriptors for Tarnoff)

8. When a person is identified only by description, it is sometimes difficult to tell whether repeated mention of the description refers to the same person or to a different person. In such cases, the answer key will represent the information in two separate objects, one of which is marked as optional.

"Last week [a spokesman for American] agreed that there was continuing progress in the negotiations with the flight attendants, but [a spokesman] said the company believed that a mediator "would help."" (In this example, the two bracketed NPs clearly refer to different persons, and the key would contain two separate objects, both obligatory.)

"As disruptions occur down the road, we need to protect ourselves and our revenues," [a company spokesman] said. ... Last week [a spokesman for American] agreed that there was continuing progress in the negotiations with the flight attendants." (For this text, the key would contain two identical objects, one of them optional, because the first bracketed NP may refer to the same person as the second bracketed NP. The actual text that inspired this example was more complicated and was not coded as described here)

"Mr. Campeau's strategy is to keep the most valuable assets, [a Wall Street source] said. "It's a sound plan," [the source] said. ... [A Wall Street source] said Mr. Campeau "was approached by another party indicating a strong interest and possibility of acquiring" Ann Taylor and Brooks Brothers. "The implication is that he tried to sell and couldn't because the price was too high," said [the source]. ... In addition to selling Allied assets, Campeau has embarked on a program to divest itself of Campeau assets in the U.S. and Canada, [a Wall Street source] said, "to give him additional liquidity to meet some of his bank requirements." In the U.S. Campeau owns property on the West Coast, including an office building in San Francisco and real estate in Silicon Valley, and some Texas real estate, [the source] said. (For this text, the key would contain three identical objects, two of them optional.)

C. For ENT_DESCRIPTOR of artifacts:

1. This slot is intended to capture information on the artifact other than its unique identifier or alias. The style for specifying a vehicle varies dependent upon whether the vehicle is one of a kind, such as ships and space shuttles, or the vehicle is mass produced with only small variations within a model line, such as automobiles, airplanes, and helicopters. The string fill for this slot is not permitted to contain the unique identifier or alias for vehicles that are one of a kind, 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 space shuttle Challenger" (descriptor is "the space shuttle")

"the Italian luxury liner Andrea Doria" (descriptor is "the Italian luxury liner")

"his square-rigger Golden Hind" (one of the alternate descriptors is "his square-rigger")

In the case of a non-unique identifier, such as the model name or an alias for the model name, the string fill for this slot is permitted to contain the identifier. A few examples of such descriptors follow with their entity names:

"the twin-engine, two-seat Tomcat" (entity name is "Tomcat")

"a chartered 757 aircraft" (entity name is "757")

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:

"airplane"
"aircraft"
"plane"
"helicopter"
"car"
"ship"

"boat"
"vehicle"
"equipment"

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 artifact in question. Taking the word "plane" as an example, the following usages of it would be regarded as insubstantial, where the expressions constitute the complete NP:

"plane" (as a bare noun, perhaps in a headline)
"the plane"
"his plane"
"that plane"
"the Boeing 737 plane" (where "plane" refers to "Boeing 737")
"XYZ Airline's Boeing 737 plane" ("plane" refers to "Boeing 737")

As a consequence of this guideline, an ARTIFACT ENTITY 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 "plane" would be regarded as substantial, and the entire phrase would be generated as ENT_DESCRIPTOR:

"a plane modeled on the Boeing 737"
"the new plane"

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., "his square-rigger" (alternate fill is "square-rigger")
b. temporal adverbials, e.g., "now the last-flying B-29" (alternate fill is "the last-flying B-29")
c. loose adjunct, e.g., a nonrestrictive relative clause or similar type of full or reduced clause, as in "the 31st off the assembly line, out of a total 694 produced to date" (alternative fill is "the 31st off the assembly line")

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 ENT_DESCRIPTOR for Commander Bates' plane that crashed outside of Nashville:

"an F-14A "Tomcat" fighter jet"
"an F-14 from Squadron 213"

9.2.2 LOCATION Object

DEFINITION: LOCATION is a task-independent object that is separate from the predefined ENTITY object. It may be defined selectively for a given scenario or relation, e.g., to provide the location of an event or ENTITY.

MINIMUM INSTANTIATION CONDITIONS: The text must supply a fill for LOCALE.

SPECIAL USAGE NOTES:
1. This object will be part of the Template Element evaluation. One or more of them may play a role in one or more Template Relation or Scenario Template tasks. In such cases, their role will be defined in the Template Relation or Scenario Template task documentation.

9.2.2.1 LOCALE Slot

DEFINITION: Specific place where an entity is located. To enable accurate, automatic scoring, only the most specific place is to be reported unless the locale is of type AIRPORT. The literal string that appears in the text appears in this slot.

MINIMUM INSTANTIATION CONDITIONS: The locale name must be specifically mentioned in the text in either noun or adjective form.

SPECIAL USAGE NOTES:
1. The "MUC-7 Reference Gazetteer" does not contain an exhaustive list of the place names that may be used

to fill the 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.
2. If the text provides only a relative locale such as “near Tokyo” or “60 miles from Tokyo”, report “Tokyo” as LOCALE.
3. If the locale is an airport, aliases are expected to be given as multiple fills rather than listing only the most complete or most specific name given in the article.

9.2.2.2 LOCALE_TYPE Slot

DEFINITION: A categorization of the place name that appears in the LOCALE slot.

MINIMUM INSTANTIATION CONDITIONS: The LOCALE slot must be filled.

SPECIAL USAGE NOTES:
1. The location categories that are to be used for LOCALE_TYPE 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
WATER -- a body of water such as the Atlantic Ocean or the Straits of Florida.
AIRPORT -- an area of land normally set aside for air traffic takeoffs and landings.
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.
2. The “MUC-7 Reference Gazetteer” uses more location categories than are to be reported in LOCALE_TYPE. The following mappings apply:
PORT in gazetteer is to be reported as CITY.
ISLAND in gazetteer is to be reported as PROVINCE.
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.

9.2.2.3 COUNTRY Slot

DEFINITION: The country or region in which LOCALE is located. A defining list of names in contained in “MUC-7 Country and Region List.” (This list contains only canonical forms. NLP system developers must define their own mappings from the “MUC-7 Reference Gazetteer” and/or other gazetteer resources to this list.)

MINIMUM INSTANTIATION CONDITIONS: To be filled if 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 LOCALE_TYPE is filled in by COUNTRY or REGION, report the name in this slot as a normalized form drawn from “MUC-7 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-7 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 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” */

6. A body of water outside the boundaries of any country or region, such as an ocean, should be reported as a string fill from the text. For example,

"the Pacific" (COUNTRY slot fill is "Pacific")
"the Mediterranean Sea" (COUNTRY slot fill is "Mediterranean Sea")
"the Mediterranean" (COUNTRY slot fill is "Mediterranean")

9.3 Known Inadequacies of the Notation

Conjoined names with elision, such as "John and Sally Smith" and "President and Mrs. Reagan", present a special problem for extraction because the text that should go in the name slot is not contiguous and other rules intercede that cause unexpected conflations of entities.

The current version of Tabula Rasa used for generating the answer key templates allows only one offset range per fill. Although, in the current evaluation, the scorer ignores offsets in scoring, they will be used in applications which display texts. So for MUC-7, we require that only contiguous strings be used wherever string fills are allowed.

In the case of "John and Sally Smith", instantiate two entities, one with ENT_NAME "John" and the other with ENT_NAME "Sally Smith." In the case of "President and Mrs. Reagan", instantiate two entities, one with ENT_DESCRIPTOR "President" and no fill for ENT_NAME and the other with the misleading fill for ENT_NAME of "Reagan" and no fill for ENT_DESCRIPTOR because "Mrs." is insubstantial.

Clearly, the issue of conjunction with elision needs to be resolved because the resulting database entries are unlikely to come to mind during the composition of database queries, especially the entry for "Mrs. Reagan" since it would conflate with the normally expected entry for her husband. Perhaps Template Relations within the database would help to clarify matters with the current rules still in effect.

APPENDIX A. Example of Template Element Objects

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

```
<ENTITY-0592-1> :=
      ENT_NAME:         "BRIDGESTONE SPORTS CO."
                        "BRIDGESTONE SPORTS"
                        "BRIDGESTON SPORTS"
      ENT_TYPE:         ORGANIZATION
      ENT_DESCRIPTOR:   "JAPANESE SPORTS GOODS MAKER"
      ENT_CATEGORY:     ORG_CO

<LOCATION-0592-1> :=
      LOCALE:           "JAPAN"
      LOCALE_TYPE:      COUNTRY
      COUNTRY:          JAPAN

<ENTITY-0592-2> :=
      ENT_NAME:         "UNION PRECISION CASTING CO."
                        "UNION PRECISION CASTING"
      ENT_TYPE:         ORGANIZATION
      ENT_DESCRIPTOR:   "A LOCALCONCERN"
                        /"CONCERN"
      ENT_CATEGORY:     ORG_CO

<LOCATION-0592-2> :=
      LOCALE:           "TAIWAN"
      LOCALE_TYPE:      COUNTRY
      COUNTRY:          TAIWAN

<ENTITY-0592-3> :=
      ENT_NAME:         "TAGA CO."
      ENT_TYPE:         ORGANIZATION
      ENT_DESCRIPTOR:   "A JAPANESE TRADING HOUSE"
                        /"A COMPANY ACTIVE IN TRADING WITH TAIWAN"
      ENT_CATEGORY:     ORG_CO

<LOCATION-0592-3> :=
      LOCALE:           "JAPAN"
      LOCALE_TYPE:      COUNTRY
      COUNTRY:          JAPAN

<ENTITY-0592-4> :=
      ENT_NAME:         "BRIDGESTONE SPORTS TAIWAN CO."
      ENT_TYPE:         ORGANIZATION
      ENT_DESCRIPTOR:   "A JOINT VENTURE"
      ENT_CATEGORY:     ORG_CO
      COMMENT:          "'A JOINT VENTURE' is the most substantive descriptor"

<LOCATION-0592-4> :=
      LOCALE:           "KAOHSIUNG"
```

```
      LOCALE_TYPE:      CITY
                        /PROVINCE
      COUNTRY:          TAIWAN
      COMMENT:          /"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."

<ENTITY-0592-5> :=
      ENT_TYPE:         ARTIFACT
      ENT_DESCRIPTOR:   "GOLF CLUBS"
                        /"GOLF CLUBS TO BE SHIPPED TO JAPAN
                        /"LUXURY CLUBS"
      ENT_CATEGORY:     ART_GROUND
      COMMENT:          "'IRON AND 'METAL WOOD' CLUBS' occurs as a substring
                        of an NP that refers to the amount of production per
                        month, not to the artifact as a general thing."

<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 20
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>
```

These are examples of fills for the Template Element task that are extracted from a text in the MUC-7 corpus. The extracted information and the text itself appear below. The ARTIFACT ENTITY assumes the rules put forth is this document that limit it to vehicles.

```
<ENTITY-9602040136-1> :=
    ENT_NAME: "NAVY"
    ENT_TYPE: ORGANIZATION
    ENT_CATEGORY: ORG_GOVT
<ENTITY-9602040136-2> :=
```

```
    ENT_NAME: "Fighter Squadron 213"
              "Squadron 213"
    ENT_TYPE: ORGANIZATION
    ENT_DESCRIPTOR: "NAVY SQUADRON"
   / "a 14-plane unit based in Miramar Naval Base near San Diego"
    ENT_CATEGORY: ORG_GOVT
<ENTITY-9602040136-3> :=
    ENT_NAME: "N.Y. Times News Service"
    ENT_TYPE: ORGANIZATION
    ENT_CATEGORY: ORG_CO
<ENTITY-9602040136-4> :=
    ENT_NAME: "Navy's Fighter Wing"
   / "Fighter Wing"
    ENT_TYPE: ORGANIZATION
    ENT_CATEGORY: ORG_GOVT
<ENTITY-9602040136-5> :=
    ENT_NAME: "Carrier Air Wing 11"
    ENT_TYPE: ORGANIZATION
    ENT_CATEGORY: ORG_GOVT
<ENTITY-9602040136-7> :=
    ENT_NAME: "Fred Kilian"
              "Kilian"
    ENT_TYPE: PERSON
    ENT_DESCRIPTOR: "NAVY SQUADRON COMMANDER"
   / "The squadron commander of the F-14 pilot in the Nashville crash that
killed five people last week"
   / "the commander"
   / "its leader"
    ENT_CATEGORY: PER_MIL
<ENTITY-9602040136-8> :=
    ENT_NAME: "JOHN O'NEIL"
    ENT_TYPE: PERSON
    ENT_CATEGORY: PER_CIV
<ENTITY-9602040136-9> :=
    ENT_NAME: "Gregg Hartung"
    ENT_TYPE: PERSON
    ENT_DESCRIPTOR: "Comdr."
   / "spokesman"
    ENT_CATEGORY: PER_MIL
<ENTITY-9602040136-10> :=
    ENT_TYPE: PERSON
    ENT_DESCRIPTOR: "Navy officer"
   / "the officer"
    ENT_CATEGORY: PER_MIL
<ENTITY-9602040136-11> :=
    ENT_NAME: "Dennis Gillespie"
    ENT_TYPE: PERSON
    ENT_DESCRIPTOR: "Capt."
   / "the commander of Carrier Air Wing 11"
    ENT_CATEGORY: PER_MIL
<ENTITY-9602040136-12> :=
    ENT_NAME: "John Stacy Bates"
   / "Bates"
    ENT_TYPE: PERSON
```

```
    ENT_DESCRIPTOR: “Lt. Comdr.”
   / “the pilot”
   / “the F-14 pilot in the Nashville crash that killed five people last week”
    ENT_CATEGORY: PER_MIL
<ENTITY-9602040136-13> :=
    ENT_NAME: “Graham Alden Higgins”
    ENT_TYPE: PERSON
    ENT_DESCRIPTOR: “Lt.”
   / “the jet’s radar operator”
    ENT_CATEGORY: PER_MIL
<ENTITY-9602040136-15> :=
    ENT_NAME: “F-14”
    ENT_TYPE: ARTIFACT
    ENT_DESCRIPTOR: “the jet”
    ENT_CATEGORY: ART_AIR
    COMMENT: “latest crash”
<ENTITY-9602040136-16> :=
    ENT_NAME: “F-14”
    ENT_TYPE: ARTIFACT
    ENT_DESCRIPTOR: “an F-14 from Squadron 213”
    ENT_CATEGORY: ART_AIR
    COMMENT: “crash in April in Pacific”
<LOCATION-9602040136-1> :=
    LOCALE: “WASHINGTON”
    LOCALE_TYPE: CITY
    COUNTRY: “United States”
<LOCATION-9602040136-2> :=
    LOCALE: “Nashville”
    LOCALE_TYPE: CITY
    COUNTRY: “United States”
<LOCATION-9602040136-3> :=
    LOCALE: “Miramar Naval Base”
    LOCALE_TYPE: AIRPORT
    COUNTRY: “United States”
<LOCATION-9602040136-4> :=
    LOCALE: “San Diego”
    LOCALE_TYPE: CITY
    COUNTRY: “United States”
<LOCATION-9602040136-5> :=
    LOCALE: “Pacific”
    LOCALE_TYPE: WATER
    COUNTRY: “Pacific Ocean”

<DOC>
<DOCID> nyt960204.0136 </DOCID>
<STORYID cat=w pri=u> A3081 </STORYID>
<SLUG fv=taf-z> BC-FIGHTER-CRASHES-300&AMP;A </SLUG>
<DATE> 02-04 </DATE>
<NWORDS> 0595 </NWORDS>
<PREAMBLE>
BC-FIGHTER-CRASHES-300&AMP;ADD-NYT
NAVY SQUADRON COMMANDER PLAGUED BY CRASHES REASSIGNED
(kd)
By JOHN O’NEIL
```

```
c.1996 N.Y. Times News Service
</PREAMBLE>
<TEXT>
<p>
   WASHINGTON  &MD;  The squadron commander of the F-14 pilot in the Nashville
crash that killed five people last week has been relieved of his command,
the Navy announced Sunday.
<p>
   Citing three accidents over the last year, the Navy decided to reassign the
commander, Fred Kilian, because of ``a loss of trust and confidence'' in his
ability to lead the squadron, said a spokesman, Comdr. Gregg Hartung.
<p>
   Fighter Squadron 213, a 14-plane unit based in Miramar Naval Base near
San Diego, had developed by far the worst safety record among the Navy's 13 F-
14 squadrons, with four crashes over the last 16 months, three after Kilian
became its leader.
<p>
   A Navy officer said that Kilian had an ``excellent reputation.''
<p>
   ``But in the Navy,'' the officer said, speaking on the condition of
anonymity, ``we hold people accountable for things that happen during the time
of their command. In this particular case, this particular squadron has an
exceptionally high accident rate  &MD;  higher than any other.''
<p>
   The officer said the decision to reassign Kilian to the Pacific
headquarters of the Navy's Fighter Wing was made Saturday by the commander
of Carrier Air Wing 11, Capt. Dennis Gillespie.
<p>
   Kilian could not be reached for comment.
<p>
   In the latest crash, an F-14 from Squadron 213 plunged to the ground
immediately after takeoff on Jan. 29, killing the pilot, Lt. Comdr. John Stacy
Bates, the jet's radar operator Lt. Graham Alden Higgins, and three
civilians in a house the plane crashed into.
<p>
   Bates had crashed an F-14 into the Pacific during a routine training flight
in April. Navy officials blamed him for causing the accident, in which no
one was hurt, by losing control, but forgave him because they felt he was
displaying the sort of aggressive flying style the Navy encourages.
</TEXT>
<TRAILER>
NYT-02-04-96 1947EST
</TRAILER>
</DOC>
```

10 Template Relation Task Object Definition

10.1 BNF

```
/* Template Relation Objects -- apply to Template Relation task; apply
selectively to Scenario Template task */

<PRODUCT_OF> :=
      ARTIFACT:          <ENTITY>-
      ORGANIZATION:      <ENTITY>-
      OBJ_STATUS:        {OPTIONAL}-
      COMMENT:           " "-
<EMPLOYEE_OF> :=
      PERSON:            <ENTITY>-
      ORGANIZATION:      <ENTITY>-
      OBJ_STATUS:        {OPTIONAL}-
      COMMENT:           " "-
<LOCATION_OF> :=
      LOCATION:          <LOCATION>-
      ORGANIZATION:      <ENTITY>-
      OBJ_STATUS:        {OPTIONAL}-
      COMMENT:           " "-
```

10.2 Fill Rules

10.2.1 PRODUCT_OF Object

DEFINITION: Relationship between an artifact and an organization entity. The artifact is a product of the organization.

MINIMUM INSTANTIATION CONDITIONS: Both the ORGANIZATION and ARTIFACT entities must appear in the text and be filled as template elements. The relation points to these objects.

10.2.2 EMPLOYEE_OF Object

DEFINITION: A relationship between a person and an organization entity. The person is an employee of the organization.

MINIMUM INSTANTIATION CONDITIONS: Both the ORGANIZATION and PERSON entities must appear in the text and be filled as template elements. The relation points to these objects.

10.2.3 LOCATION_OF Object

DEFINITION: A relationship between a LOCATION and an ORGANIZATION.

MINIMUM INSTANTIATION CONDITIONS: Both the ORGANIZATION and LOCATION must appear in the text and be filled as template elements. The relation points to these objects.

APPENDIX A. Example of Template Relation Objects

These are examples of fills for the Template Relation task that are extracted from a text in the MUC-5 corpus. The extracted information and the text itself appear below. The ARTIFACT ENTITY assumes a scenario-specific IE task that includes sports equipment such as golf clubs (and excludes such things as “golf club parts”) and specifies where they are used in the ENT_CATEGORY slot.

```
<ENTITY-0592-1> :=
      ENT_NAME:         "BRIDGESTONE SPORTS CO."
                        "BRIDGESTONE SPORTS"
                        "BRIDGESTON SPORTS"
      ENT_TYPE:         ORGANIZATION
      ENT_DESCRIPTOR:   "JAPANESE SPORTS GOODS MAKER"
      ENT_CATEGORY:     ORG_CO

<LOCATION-0592-1> :=
      LOCALE:           "JAPAN"
      LOCALE_TYPE:      COUNTRY
      COUNTRY:          JAPAN

<ENTITY-0592-2> :=
      ENT_NAME:         "UNION PRECISION CASTING CO."
                        "UNION PRECISION CASTING"
      ENT_TYPE:         ORGANIZATION
      ENT_DESCRIPTOR:   "A LOCALCONCERN"
                        /"CONCERN"
      ENT_CATEGORY:     ORG_CO

<LOCATION-0592-2> :=
      LOCALE:           "TAIWAN"
      LOCALE_TYPE:      COUNTRY
      COUNTRY:          TAIWAN

<ENTITY-0592-3> :=
      ENT_NAME:         "TAGA CO."
      ENT_TYPE:         ORGANIZATION
      ENT_DESCRIPTOR:   "A JAPANESE TRADING HOUSE"
                        /"A COMPANY ACTIVE IN TRADING WITH TAIWAN"
      ENT_CATEGORY:     ORG_CO

<ENTITY-0592-4> :=
      ENT_NAME:         "BRIDGESTONE SPORTS TAIWAN CO."
      ENT_TYPE:         ORGANIZATION
      ENT_DESCRIPTOR:   "A JOINT VENTURE"
      ENT_CATEGORY:     ORG_CO
      COMMENT:          "'A JOINT VENTURE' is the most substantive descriptor"

<LOCATION-0592-3> :=
      LOCALE:           "KAOHSIUNG"
      LOCALE_TYPE:      CITY
                        /PROVINCE
      COUNTRY:          TAIWAN
      COMMENT:          /"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.”

<ENTITY-0592-5> :=
      ENT_TYPE:         ARTIFACT
      ENT_DESCRIPTOR:   ”GOLF CLUBS”
                        /”GOLF CLUBS TO BE SHIPPED TO JAPAN
                        /”LUXURY CLUBS”
      ENT_CATEGORY:     ART_GROUND
      COMMENT:          “‘IRON AND ‘METAL WOOD’ CLUBS’ occurs as a substring
                        of an NP that refers to the amount of production per
                        month, not to the artifact as a general thing.”

<PRODUCT_OF-0592-1> :=
      ARTIFACT:         <ENTITY-0592-5>
      ORGANIZATION:     <ENTITY-0592-4>

<LOCATION_OF-0592-1> :=
      LOCATION:         <LOCATION-0592-3>
      ORGANIZATION:     <ENTITY-0592-4>

<LOCATION_OF-0592-2> :=
      LOCATION:         <LOCATION-0592-2>
      ORGANIZATION:     <ENTITY-0592-2>

<LOCATION_OF-0592-3> :=
      LOCATION:         <LOCATION-0592-1>
      ORGANIZATION:     <ENTITY-0592-3>
      COMMENT:          “A JAPANESE TRADING HOUSE;TAGA CO., A COMPANY ACTIVE
                        IN TRADING WITH TAIWAN,“

<LOCATION_OF-0592-4> :=
      LOCATION:         <LOCATION-0592-1>
      ORGANIZATION:     <ENTITY-0592-1>

<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 20
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>
```

These are examples of fills for the Template Relation task that are extracted from a text in the MUC-7 corpus. The extracted information and the text itself appear below. The ARTIFACT ENTITY assumes the rules put forth is this document that limit it to vehicles.

```
<TEMPLATE-9602040136-1> :=
    DOC_NR: "9602040136"
<ENTITY-9602040136-1> :=
    ENT_NAME: "NAVY"
    ENT_TYPE: ORGANIZATION
    ENT_CATEGORY: ORG_GOVT
<ENTITY-9602040136-2> :=
    ENT_NAME: "Fighter Squadron 213"
              "Squadron 213"
    ENT_TYPE: ORGANIZATION
    ENT_DESCRIPTOR: "NAVY SQUADRON"
   / "a 14-plane unit based in Miramar Naval Base near San Diego"
    ENT_CATEGORY: ORG_GOVT
<ENTITY-9602040136-3> :=
    ENT_NAME: "N.Y. Times News Service"
    ENT_TYPE: ORGANIZATION
    ENT_CATEGORY: ORG_CO
<ENTITY-9602040136-4> :=
    ENT_NAME: "Navy's Fighter Wing"
            / "Fighter Wing"
    ENT_TYPE: ORGANIZATION
    ENT_CATEGORY: ORG_GOVT
<ENTITY-9602040136-5> :=
    ENT_NAME: "Carrier Air Wing 11"
    ENT_TYPE: ORGANIZATION
    ENT_CATEGORY: ORG_GOVT
<ENTITY-9602040136-6> :=
    ENT_NAME: "Air National Guard"
    ENT_TYPE: ORGANIZATION
    ENT_CATEGORY: ORG_GOVT
<ENTITY-9602040136-7> :=
    ENT_NAME: "Fred Kilian"
              "Kilian"
    ENT_TYPE: PERSON
    ENT_DESCRIPTOR: "NAVY SQUADRON COMMANDER"
   / "The squadron commander of the F-14 pilot in the Nashville crash that
killed five people last week"
   / "the commander"
   / "its leader"
   / "leader"
    ENT_CATEGORY: PER_MIL
<ENTITY-9602040136-8> :=
```

```
    ENT_NAME: "JOHN O'NEIL"
    ENT_TYPE: PERSON
    ENT_CATEGORY: PER_CIV
<ENTITY-9602040136-9> :=
    ENT_NAME: "Gregg Hartung"
              "Hartung"
    ENT_TYPE: PERSON
    ENT_DESCRIPTOR: "Comdr."
   / "spokesman"
   / "Navy spokesman"
    ENT_CATEGORY: PER_MIL
<ENTITY-9602040136-10> :=
    ENT_TYPE: PERSON
    ENT_DESCRIPTOR: "Navy officer"
   / "the officer"
    ENT_CATEGORY: PER_MIL
<ENTITY-9602040136-11> :=
    ENT_NAME: "Dennis Gillespie"
    ENT_TYPE: PERSON
    ENT_DESCRIPTOR: "Capt."
   / "the commander of Carrier Air Wing 11"
    ENT_CATEGORY: PER_MIL
<ENTITY-9602040136-12> :=
    ENT_NAME: "John Stacy Bates"
              "Bates"
    ENT_TYPE: PERSON
    ENT_DESCRIPTOR: "Lt. Comdr."
   / "the pilot"
   / "the F-14 pilot in the Nashville crash that killed five people last week"
    ENT_CATEGORY: PER_MIL
<ENTITY-9602040136-13> :=
    ENT_NAME: "Graham Alden Higgins"
    ENT_TYPE: PERSON
    ENT_DESCRIPTOR: "Lt."
   / "the jet's radar operator"
    ENT_CATEGORY: PER_MIL
<ENTITY-9602040136-14> :=
    ENT_NAME: "Kara S. Hultgreen"
    ENT_TYPE: PERSON
    ENT_DESCRIPTOR: "Lt."
   / "one of the Navy's first female fighter pilots"
    ENT_CATEGORY: PER_MIL
<ENTITY-9602040136-15> :=
    ENT_NAME: "F-14"
    ENT_TYPE: ARTIFACT
    ENT_DESCRIPTOR: "the jet"
   / "an F-14 from Squadron 213"
    ENT_CATEGORY: ART_AIR
    COMMENT: "latest crash"
<ENTITY-9602040136-16> :=
    ENT_NAME: "F-14"
    ENT_TYPE: ARTIFACT
    ENT_CATEGORY: ART_AIR
    COMMENT: "crash in April in Pacific"
```

```
<ENTITY-9602040136-17> :=
    ENT_NAME: "F-14"
    ENT_TYPE: ARTIFACT
    ENT_DESCRIPTOR: "an F-14 from Squadron 213"
    ENT_CATEGORY: ART_AIR
<ENTITY-9602040136-18> :=
    ENT_TYPE: ARTIFACT
    ENT_DESCRIPTOR: "an aircraft carrier"
    ENT_CATEGORY: ART_WATER
    OBJ_STATUS: OPTIONAL
<LOCATION-9602040136-1> :=
    LOCALE: "WASHINGTON"
    LOCALE_TYPE: CITY
    COUNTRY: "United States"
<LOCATION-9602040136-2> :=
    LOCALE: "Nashville"
    LOCALE_TYPE: CITY
    COUNTRY: "United States"
<LOCATION-9602040136-3> :=
    LOCALE: "Miramar Naval Base"
    LOCALE_TYPE: AIRPORT
    COUNTRY: "United States"
<LOCATION-9602040136-4> :=
    LOCALE: "San Diego"
    LOCALE_TYPE: CITY
    COUNTRY: "United States"
<LOCATION-9602040136-5> :=
    LOCALE: "Pacific"
    LOCALE_TYPE: WATER
    COUNTRY: "Pacific Ocean"
<LOCATION-9602040136-6> :=
    LOCALE: "Berry Field"
    LOCALE_TYPE: AIRPORT
    COUNTRY: "United States"
<LOCATION-9602040136-7> :=
    LOCALE: "Nashville International Airport"
    LOCALE_TYPE: AIRPORT
    COUNTRY: "United States"
<LOCATION-9602040136-8> :=
    LOCALE: "Southern California"
    LOCALE_TYPE: PROVINCE
    COUNTRY: "United States"
<EMPLOYEE_OF-9602040136-1> :=
    PERSON: <ENTITY-9602040136-7>
    ORGANIZATION: <ENTITY-9602040136-1>
<EMPLOYEE_OF-9602040136-2> :=
    PERSON: <ENTITY-9602040136-8>
    ORGANIZATION: <ENTITY-9602040136-3>
<EMPLOYEE_OF-9602040136-3> :=
    PERSON: <ENTITY-9602040136-9>
    ORGANIZATION: <ENTITY-9602040136-1>
<EMPLOYEE_OF-9602040136-4> :=
    PERSON: <ENTITY-9602040136-10>
    ORGANIZATION: <ENTITY-9602040136-1>
```

```
<EMPLOYEE_OF-9602040136-5> :=
    PERSON: <ENTITY-9602040136-11>
    ORGANIZATION: <ENTITY-9602040136-1>
<EMPLOYEE_OF-9602040136-6> :=
    PERSON: <ENTITY-9602040136-12>
    ORGANIZATION: <ENTITY-9602040136-1>
<EMPLOYEE_OF-9602040136-7> :=
    PERSON: <ENTITY-9602040136-13>
    ORGANIZATION: <ENTITY-9602040136-1>
<EMPLOYEE_OF-9602040136-8> :=
    PERSON: <ENTITY-9602040136-14>
    ORGANIZATION: <ENTITY-9602040136-1>
<LOCATION_OF-9602040136-1> :=
    LOCATION: <LOCATION-9602040136-1>
    ORGANIZATION: <ENTITY-9602040136-1>
    OBJ_STATUS: OPTIONAL
    COMMENT: “implied”
<LOCATION_OF-9602040136-2> :=
    LOCATION: <LOCATION-9602040136-2>
    ORGANIZATION: <ENTITY-9602040136-1>
    OBJ_STATUS: OPTIONAL
    COMMENT: “unclear”
<LOCATION_OF-9602040136-3> :=
    LOCATION: <LOCATION-9602040136-3>
    ORGANIZATION: <ENTITY-9602040136-2>
<LOCATION_OF-9602040136-4> :=
    LOCATION: <LOCATION-9602040136-4>
    ORGANIZATION: <ENTITY-9602040136-1>
<LOCATION_OF-9602040136-5> :=
    LOCATION: <LOCATION-9602040136-6>
    ORGANIZATION: <ENTITY-9602040136-6>
```

```
<DOC>
<DOCID> nyt960204.0136 </DOCID>
<STORYID cat=w pri=u> A3081 </STORYID>
<SLUG fv=taf-z> BC-FIGHTER-CRASHES-300&AMP;A </SLUG>
<DATE> 02-04 </DATE>
<NWORDS> 0595 </NWORDS>
<PREAMBLE>
BC-FIGHTER-CRASHES-300&AMP;ADD-NYT
NAVY SQUADRON COMMANDER PLAGUED BY CRASHES REASSIGNED
(kd)
By JOHN O’NEIL
c.1996 N.Y. Times News Service
</PREAMBLE>
<TEXT>
<p>
   WASHINGTON  &MD;  The squadron commander of the F-14 pilot in the Nashville
crash that killed five people last week has been relieved of his command,
the Navy announced Sunday.
<p>
   Citing three accidents over the last year, the Navy decided to reassign the
commander, Fred Kilian, because of ``a loss of trust and confidence’’ in his
```

```
ability to lead the squadron, said a spokesman, Comdr. Gregg Hartung.
<p>
   Fighter Squadron 213, a 14-plane unit based in Miramar Naval Base near
San Diego, had developed by far the worst safety record among the Navy's 13 F-
14 squadrons, with four crashes over the last 16 months, three after Kilian
became its leader.
<p>
   A Navy officer said that Kilian had an ``excellent reputation.''
<p>
   ``But in the Navy,'' the officer said, speaking on the condition of
anonymity, ``we hold people accountable for things that happen during the time
of their command. In this particular case, this particular squadron has an
exceptionally high accident rate  &MD;  higher than any other.''
<p>
   The officer said the decision to reassign Kilian to the Pacific
headquarters of the Navy's Fighter Wing was made Saturday by the commander
of Carrier Air Wing 11, Capt. Dennis Gillespie.
<p>
   Kilian could not be reached for comment.
<p>
   In the latest crash, an F-14 from Squadron 213 plunged to the ground
immediately after takeoff on Jan. 29, killing the pilot, Lt. Comdr. John Stacy
Bates, the jet's radar operator Lt. Graham Alden Higgins, and three
civilians in a house the plane crashed into.
<p>
   Bates had crashed an F-14 into the Pacific during a routine training flight
in April. Navy officials blamed him for causing the accident, in which no
one was hurt, by losing control, but forgave him because they felt he was
displaying the sort of aggressive flying style the Navy encourages.
</TEXT>
<TRAILER>
NYT-02-04-96 1947EST
</TRAILER>
</DOC>
```

11 Scenario Template Task Object Definition

This scenario concerns aircraft crash or accident incident reports and updates. The template captures information about the aircraft involved in the accident or crash, and the location and the date of the accident or crash, the number of victims in the accident, and information on the status of the investigation with respect to the cause of the crash. It also captures information of reports about the incident within the article. The event object captures flight information, the date and location of the accident or crash, causal information, and reporting information. The relational and low-level objects capture flight information on aircraft, including departure and arrival information, specific causal information, and information on individual reports within the context of the article. A relevant article refers to an aircraft crash or accident that has actually occurred and must minimally identify the aircraft and its owner (an organization or person), date and location of crash or accident, aircraft type, number of victims, and the cause of the accident.The location and date of the accident or crash are either specified directly or can be inferred from information in the article.

11.1 Template BNF

```
<TEMPLATE> :=
      DOC_NR:      "NUMBER" ^
      CONTENT:     <CRASH_EVENT> *
      COMMENT:     "COMMENT"

<CRASH_EVENT> :=
      FLIGHT_INFO:<FLIGHT_INFO> ^
      AIRCRAFT_TYPE:{ MILITARY, COMMUTER, COMMERCIAL, PRIVATE }  ^
      CRASH_DATE:      <TIME> ^
      CRASH_SITE: <LOCATION> ^
      CAUSE_INFO: <CAUSE_INFO> ^
      VICTIMS_NUM:[TEXT] ^
      REPORT_EVENT:<REPORT_EVENT> *
      COMMENT:     "COMMENT"

<FLIGHT_INFO> :=
      AIRCRAFT:    <ENTITY> ^
      AIRLINE:     <ENTITY> ^
      MANUFACTURER:<ENTITY>
      DESTINATION:<LOCATION>
      DEPARTURE_POINT:<LOCATION>
      DEPARTURE_DATE:<TIME>
      COMMENT:     "COMMENT"

<CAUSE_INFO> :=
      CAUSE_TYPE: { EQUIPMENT_FAILURE, WEATHER, HUMAN_ERROR, BOMB, CARGO,
                  MISSILE, OTHER, UNKNOWN } +
      LIKELIHOOD: { YES, UNCLEAR } ^
      COMMENT:     "COMMENT"

<REPORT_EVENT> :=
      REPORTER:    <ENTITY> ^
      AFFILIATION:<ENTITY>
      REPORTER_ROLE:{ SPOKESPERSON, PASSENGER, WITNESS, INVESTIGATOR,
                   RELATIVE }  ^
      REPORT_DATE:<TIME>
      REPORT_LOCATION:<LOCATION>
      COMMENT:     "COMMENT"
```

```
<TIME> :=
      START:       DATE-EXP
      END:         DATE-EXP
      DESCRIPTOR: "DESCRIPTOR" ^
      OBJ_STATUS: { OPTIONAL }
      COMMENT:     "COMMENT"

DATE-EXP ::([[01-31]]|{EA, MD, LT, EO, BO})[[01-12]][[1000-9999]YYYY]
                  | {EA, MD, LT, EO, BO}
                    {FA, WI, SP, SU, 1Q, 2Q, 3Q, 4Q, 1F, 2F, 3F, 4F, FY}
                    [[1000-9999]YYYY]
                  | {EA, MD, LT, EO, BO, FA, WI, SP, SU, 1Q, 2Q, 3Q, 4Q,
                       1F, 2F, 3F, 4F, FY}
                    [[1000-9999]YYYY]
                  | [[01-12]][[1000-9999]YYYY]
                  | [[1000-9999]]
                  | UNCLEAR
```

11.2 Fill Rules

Note that each object defined by the template BNF contains COMMENT and OBJ_STATUS slots. These slots are reserved for use in the answer-key templates; they are not to be generated by the extraction systems. See *Information Extraction Task Definition* document for further information.

11.2.1 TEMPLATE Object

DEFINITION: Top-level object. See *Information Extraction Task Definition* for further information.

11.2.1.1 DOC_NR Slot

DEFINITION: Document identifier. See *Information Extraction Task Definition* for further information.

11.2.1.2 CONTENT Slot

DEFINITION: Pointer to object that captures info on the aircraft accident or crash. See *Information Extraction Task Definition* for general information.

MINIMUM INSTANTIATION CONDITIONS: The article must identify or refer to an aircraft accident or crash, must identify the aircraft, the owner of the aircraft (either a person or organization), the crash location and date, the number of victims, and causal information concerning the crash. To be relevant, it must be the case that the accident or crash actually occurred and that a specific date and location are specified or inferred. More than one crash event may be mentioned in an article. If there are multiple events instantiated, additional events must meet the same minimum instantiation conditions.

SPECIAL USAGE NOTES:
1. Relevant types of crashes or accidents are limited to those that have actually occurred. Hypothetical discussions of aircraft accidents or crashes ("what if...") are not relevant.
2. More than one crash event may be mentioned in an article. For a second crash event to be relevant, it must be related in cause, aircraft type, or flight info information to the first relevant crash event.
3. There must be one fully instantiated CRASH_EVENT object for the article to be relevant. Additional CRASH_EVENT objects must meet the same minimum instantiation conditions.For example, consider an article where Flight 800 is being discussed and Pan Am 103 is mentioned as well. If both the TWA Flight 800 CRASH_EVENT object and the Pan Am 103 CRASH_EVENT object meet minimal instantiation criteria (i.e. aircraft, airline, date and location of crash or accident, aircraft type, victims, and cause are specified), there will be two CRASH_EVENT objects instantiated.
4. In some articles, mention of a crash event may not meet minimum instantiation conditions, e.g. an article discussing Flight 800 4 months after the crash might refer merely to Flight 800 and not provide crash date and/ or location information since that information is part of the general knowledge of the reader. For the sake of consistency, there will be no CRASH_EVENT object generated for the Flight 800 crash.
5. Mention of a crash will not necessarily instantiate the CRASH_EVENT. For example, there may be mention

of the "Pan Am crash in Lockerbie, Scotland in 1988". If no specific aircraft is mentioned, then the CRASH_EVENT object will not be instantiated

11.2.2 CRASH_EVENT Object.

DEFINITION: The aircraft crash or accident event.

MINIMUM INSTANTIATION CONDITIONS: See section 3.1. 2 above.

SPECIAL USAGE NOTES:
1. Instantiate separate CRASH_EVENTS for each aircraft accident or crash, subject to instantiation conditions.
2. Instantiate separate CRASH_EVENT objects for each accident or crash an aircraft may participate in. An aircraft may be involved in more than one CRASH_EVENT.

11.2.2.1 FLIGHT_INFO Slot

DEFINITION: Pointer to object carrying information about aircraft, airline, manufacturer, departure date, location and destination

MINIMUM INSTANTIATION CONDITIONS: This slot must always be filled

11.2.2.2 AIRCRAFT_TYPE Slot

DEFINITION: A general categorization of the type of aircraft involved in the crash.

MINIMUM INSTANTIATION CONDITIONS: This slot must always be filled.

SPECIAL USAGE NOTES:
1. The categories that are to be used for this slot are defined as follows:
MILITARY -- A military aircraft.
COMMUTER -- A commuter aircraft. Typically a smaller aircraft used regularly for short hops between cities.
COMMERCIAL -- A commercial aircraft.; can be either passenger or cargo aircraft.
PRIVATE -- An aircraft carrying only several passengers. This type of aircraft may be owned by an individual or by an organization. Both airplanes and helicopters may fall under this category.

11.2.2.3 CRASH_DATE slot

DEFINITION: Pointer to an object indicating the date of the crash or accident.

MINIMUM INSTANTIATION CONDITIONS: This slot must always be filled.

SPECIAL USAGE NOTES: See section 3.6 below and time-guidelines-supplement-07aug97.
1. If two mentions of dates for the crash or accident are given, the most specific date should be extracted. For example, if the choice is between extracting "last week" or "January 17" as the date of the crash or accident, choose "January 17".

11.2.2.4 CRASH_SITE Slot

DEFINITION: Pointer to an object indicating the location of the crash or accident

MINIMUM INSTANTIATION CONDITIONS: This slot must always be filled.

SPECIAL USAGE NOTES:
1. The most specific location information should be used. For example, if the crash occurred "in the Atlantic Ocean near Long Island", then a pointer to the LOCATION object for LONG ISLAND should be used.

11.2.2.5 CAUSE_INFO Slot

DEFINITION: Pointer to an object holding information about the cause of the accident or crash.

MINIMUM INSTANTIATION CONDITIONS: This slot must be filled.

SPECIAL USAGE NOTES: (See section 11.2.4 for further information.)

11.2.2.6 VICTIMS_NUM Slot

DEFINITION: Specification of the number of victims in the crash. This includes the sum of the number of dead and injured if both are given.

MINIMUM INSTANTIATION CONDITIONS: This slot must be filled.

SPECIAL USAGE NOTES: This slot is to be filled by a numerical expression (0 is a possible fill).

11.2.2.7 REPORT_EVENT Slot

DEFINITION: A pointer to the object that captures information on the person reporting on the incident or updating the incident report within the context of the article. It does not contain information on the reporter who wrote the article unless that reporter is mentioned within the body of the article.

MINIMUM INSTANTIATION CONDITIONS: The article must identify the reporter by name or descriptor.

SPECIAL USAGE NOTES:
1. Instantiate separate REPORT_EVENT objects for each report event mentioned in the article.

11.2.3 FLIGHT_INFO Object

DEFINITION: This object relates flight information about an aircraft that was involved in an accident or crash.

MINIMUM INSTANTIATION CONDITIONS: See section 3.2.1.

11.2.3.1 AIRCRAFT Slot

DEFINITION: Pointer to the object that identifies and names the aircraft involved in the accident or crash.

MINIMUM INSTANTIATION CONDITIONS: This slot must be filled. The name of an aircraft or a descriptor must be given.

SPECIAL USAGE NOTES:
1. Text must directly identify the aircraft model or series. For example, the aircraft model or series would be things like LC-1011, DC-9; 737. A pointer to an ENTITY with only the ENT_DESCRIPTOR “TWA flight” and no ENT_NAME would not be appropriate.

11.2.3.2 AIRLINE Slot

DEFINITION: Pointer to the object that identifies the owner of the aircraft. This will either be an ENTITY of type ORGANIZATION or of type PERSON.

MINIMUM INSTANTIATION CONDITIONS: This slot must be filled.

SPECIAL USAGE NOTES:
1. For COMMERCIAL aircraft, AIRLINE will be filled by a pointer to an ENTITY of type ORGANIZATION.
2. For PRIVATE aircraft, AIRLINE will be filled by a pointer to an ENTITY of either type ORGANIZATION or PERSON.
3. For MILITARY aircraft, AIRLINE will be filled by a pointer to an ENTITY object of type ORGANIZATION, with the category ORG_GOVT. This includes, for example, a branch of the military (such as the “Navy”).

11.2.3.3 MANUFACTURER Slot

DEFINITION: Identification of the company that manufactured the aircraft.

MINIMUM INSTANTIATION CONDITIONS: This slot may be empty.

SPECIAL USAGE NOTES:
1. The name or description of the manufacturing organization must be given.
2. Relevant types of organizations for the MANUFACTURER slot are limited to those that are categorized as ENT_TYPE ORGANIZATION where ENT_CATEGORY is ORG_CO.
3. If the aircraft is a military aircraft, the manufacturer must be identified independently of the military in order for this slot to be filled. For example, for the object referred to by “a military version of the Boeing 737”, the MANUFACTURER is “Boeing” and the AIRLINE is “Air Force”.

11.2.3.4 DESTINATION Slot

DEFINITION: Pointer to an object indicating the location to which the aircraft was headed when the accident or crash occurred.

MINIMUM INSTANTIATION CONDITIONS: This slot may be empty.

SPECIAL USAGE NOTES:
1. Use the most specific LOCATION information available. Typically this would be a city name or an airport. If both are available, choose the airport.

11.2.3.5 DEPARTURE_POINT Slot

DEFINITION: Pointer to an object indicating the location from which the aircraft departed.

MINIMUM INSTANTIATION CONDITIONS: This slot may be empty.

SPECIAL USAGE NOTES:
1. Use the most specific LOCATION information available. Typically this would be a city name or an airport. If both are available, choose the airport.

11.2.3.6 DEPARTURE_DATE Slot

DEFINITION: Pointer to an object indicating the date at which the aircraft departed the DEPARTURE_POINT.

MINIMUM INSTANTIATION CONDITIONS: This slot may be empty.

SPECIAL USAGE NOTES:
1. A completely specified date must be used, i.e. DDMMYYYY.
2. Unless otherwise indicated, the DEPARTURE_DATE will be the same as the CRASH_DATE, as long as the CRASH_DATE has also been full specified.

11.2.4 CAUSE_INFO Object

DEFINITION: Information about the cause or hypothesized cause of the aircraft accident or crash.

MINIMUM INSTANTIATION CONDITIONS: See section 3.2.5.

11.2.4.1 CAUSE_TYPE Slot

DEFINITION: Identification of the cause of the crash or accident with respect to the CRASH_EVENT as of the date of the article, i.e. information on what the current thought is on the cause of the accident at the time the news appeared. Multiple cause types may be specified.

MINIMUM INSTANTIATION CONDITIONS: This slot must always be filled; the default fill is UNKNOWN.

SPECIAL USAGE NOTES:
1. The categories that are to used for this slot are defined as follows:
EQUIPMENT_FAILURE -- some problem with the aircraft's equipment caused the crash or accident.
WEATHER -- weather conditions led to aircraft crash or accident.
HUMAN_ERROR -- human error precipitated crash or accident. This may include errors by Air Traffic Controllers, ground crew, mechanics, service people, airline or airport personnel.
BOMB -- bomb on aircraft caused an explosion that led to crash or accident.
CARGO -- item on aircraft [not a bomb] exploded or caught fire, causing crash or accident.
MISSILE -- a missile sent by friendly or enemy forces caused the crash or accident.
OTHER -- some cause other then those specified above caused the crash
UNKNOWN -- article specifies that the cause of the crash is unknown.
2. Hypotheses as to the cause of the accident or crash are adequate.
3. Causal information must be categorized and a measure of likelihood assigned. A fill of YES will be used when all of the CAUSES specified by the set are true and UNKNOWN is no longer a member of the set. UNCLEAR will be used otherwise.
4. If a CAUSE_TYPE has been eliminated in the investigation of the crash or accident, then it will not appear in the set CAUSE_TYPE.

11.2.4.2 LIKELIHOOD Slot

DEFINITON: Identification of the status of the CAUSE_TYPES with respect to the date of the article.

MINIMUM INSTANTIATION CONDITIONS: This slot must always be filled The default is UNCLEAR.

SPECIAL USAGE NOTES:
1. The categories that are to be used for this slot are defined as follows:
YES -- the CAUSE_TYPE(s) are the actual cause(s) of the aircraft accident or crash.
UNCLEAR -- The causes specified are still under investigation.
2. If more than one CAUSE_TYPE is specified, then the LIKELIHOOD value refers to the set.
3. The only time YES will appear as LIKELIHOOD will be when the cause(s) of the accident have been identified.
4. If CAUSE_TYPE is UNKNOWN, then LIKELIHOOD is UNCLEAR.

11.2.5 REPORT_EVENT Object

DEFINITON: This object relates information concerned with a report of the incident mentioned in the article. It includes the reporter, affiliation of the reporter, role of the reporter, date and location of the report.

MINIMUM INSTANTIATION CONDITIONS: See Section 3.2.7

SPECIAL USAGE NOTES:
1. The REPORT_EVENT concerns a report of the crash incident mentioned in the article. It does not refer to the reporting of the article itself. So REPORT_EVENT does not cover meta-information about the article (e.g. reporter, dateline, location of newspaper, etc.).
2. There may be multiple REPORT_EVENTS.
3. If the REPORTER does not fall into any of the categories specified by REPORTER_ROLE, then do not instantiate the REPORT_EVENT. For example, if the commentator is a public servant who is merely commenting on the crash or accident investigation, then do not instantiate the REPORT_EVENT.

11.2.5.1 REPORTER Slot

DEFINITION: Pointer to an ENTITY of type PERSON or ORGANIZATION that captures the information on the reporter of the CRASH_EVENT incident within the context of the article. It is not the newspaper reporter who wrote the article.

MININIMUM INSTANTIATION CONDITIONS: The name of a person, organization, or descriptor must be given.

SPECIAL USAGE NOTES:
1. Reporter must provide information about the incident, the investigation of the incident, or cause of the incident.
2. Reporter does not merely pay tribute to victim, discuss the consequences of the victim's death, or make public statements concerning the victims or incident.

11.2.5.2 AFFILIATION Slot

DEFINITION: Pointer to an ENTITY object of type ORGANIZATION that captures information on the affiliation of the REPORTER.

MINIMUM INSTANTIATION CONDITIONS: This slot may be empty.

SPECIAL USAGE NOTES:
1. ENT_NAME of the ORGANIZATION ENTITY must be specified. A descriptor alone is not sufficient.
2. AFFILIATION may not reflect the REPORTER_ROLE. For example, a bartender at "Joe's Bar and Grill" may be a WITNESS to the accident.
3. If a REPORTER is of type ORGANIZATION, do not fill the AFFILIATION slot with the redundant information.

11.2.5.3 REPORTER_ROLE Slot

DEFINITION: The identification of the ROLE of the REPORTER within the context of this article.

MINIMUM INSTANTIATION CONDITIONS: This slot must always be filled if the REPORTER is filled.

SPECIAL USAGE NOTES:
1. The categories that are used for this slot are defined as follows:
SPOKESPERSON -- the REPORTER is a spokesperson for an organization.
PASSENGER -- the REPORTER was a passenger or crew member on the aircraft.
WITNESS -- the REPORTER witnessed the aircraft accident or crash.
INVESTIGATOR -- the REPORTER is involved in the investigation of the cause of the crash or accident.
RELATIVE -- the REPORTER is a relative of a passenger or crew member on the aircraft.
2. INVESTIGATOR includes experts and consultants as well as people participating in the search and rescue and salvage operations.
3. The role of WITNESS takes precedence over INVESTIGATOR. If a person has WITNESSed the accident or crash and also participates in the rescue operations, identify the REPORTER as a WITNESS.

11.2.5.4 REPORT_DATE Slot

DEFINITION: A pointer to an object which indicates the date on which the reporter made the report.

MINIMUM INSTANTIATION CONDITIONS: The default is the date of the article.

11.2.5.5 REPORT_LOCATION Slot

DEFINITION: A pointer to an object that indicates the location from which the report was made.

MINIMUM INSTANTIATION CONDITIONS: This slot may be empty.

SPECIAL USAGE NOTES:
1. The most specific location information should be used.

11.2.6 TIME Object

DEFINITION: An absolute or relative date expressed as a date range.

MINIMUM INSTANTIATION CONDITIONS: "See "A Revised Template Description for Time (v3)" prepared by James Pustejovsky and Federica Busa.

SPECIAL USAGE NOTES:
1. Interpretation of the template element may require calculation from the reference date of the article.
2. DESCRIPTOR is a required slot if there is a time expression.
3. Time of day information will not be extracted for MUC-7. However, if two events occur at different times during the day, e.g. Saturday morning vs. Saturday evening, different TIME objects will be instantiated.
4. Calendar information about days of the week will be recognized. For example, last Saturday, if DD=8/8/97, would have the DATE-EXP = 02081997.
5. The START and END values represent the longest possible duration of the interval, e.g., START 290687 starts just after midnight on the 29th and END 010787 ends at midnight on the 1st. The times need not be specified in the date expression. For example, "in the next 2 days" means that the START value is today, and the END value is today+2.
6. Weeks will not be calculated. It is usually unclear, what is meant by expressions like: next week, this week, last week, in two weeks, i.e. whether the week is Sun-Sat or is based on the date of the article. UNCLEAR will be used as required. For example, "this week" [where DD is "07/15/87"] would be expressed as

```
TIME:        START 15071987
             END: UNCLEAR
             DESCRIPTOR: "this week"
```

7. UNCLEAR will be used if the START or END points are not specified fully or cannot be inferred from other time information. This will be used for such expressions as "before June 5, 1987", where the START point is UNCLEAR but the END point is "June 5, 1987."
8. Intermittent time expressions will be expressed by the most recent occurrence. For example, "every January 28th" only the most recent will be specified:

```
TIME:        START: 28011997
             END: 28011997
```

If an intermittent time expression has more specific interval information, that will be used. For example, if the time expression were "every January 28th for the last 10 years", the object would look like:

TIME:	START:	28011987
	END:	28011997