MUC-7 Information Extraction Task Definition

Version 4.2

6 March 1998

Nancy Chinchor (chinchor@gso.saic.com)

1 Brief Definition of Information Extraction Task . . . 4
2 Motivation and Timing . . . 4
2.1 Domain-Independent Information Extraction . . . 4
2.2 Domain-Dependent Information Extraction . . . 4
3 General Description of IE Tasks . . . 4
3.1 Template Element . . . 4
3.2 Template Relation . . . 5
3.3 Scenario Template . . . 5
4 Evaluation Stages . . . 6
5 Levels of Template Structure . . . 6
6 Fill Format . . . 7
6.1 Slot Types . . . 7
6.2 Object Identifiers . . . 7
6.3 Notation Reserved for Use in Answer Keys . . . 7
7 Document Sections for Information Extraction . . . 7
8 Description of Task Defining Documents . . . 7
9 Template Element Task Object Definition . . . 9
9.1 BNF . . . 9
9.2 Fill Rules . . . 9
9.2.1 ENTITY Object . . . 9
9.2.1.1 ENT_NAME Slot . . . 9
9.2.1.2 ENT_TYPE Slot . . . 10
9.2.1.3 ENT_CATEGORY Slot . . . 10
9.2.1.4 ENT_DESCRIPTOR Slot . . . 10
9.2.2 LOCATION Object . . . 14
9.2.2.1 LOCALE Slot . . . 14
9.2.2.2 LOCALE_TYPE Slot . . . 15
9.2.2.3 COUNTRY Slot . . . 15
9.3 Known Inadequacies of the Notation . . . 16
10 Template Relation Task Object Definition . . . 22
10.1 BNF . . . 22
10.2 Fill Rules . . . 22
10.2.1 General Template Relation Object . . . 22
10.2.2 PRODUCT_OF Object . . . 22
10.2.3 EMPLOYEE_OF Object . . . 23
10.2.4 LOCATION_OF Object . . . 25
10.3 Scoring Note . . . 26
11 Scenario Template Task Object Definition . . . 34
11.1 Template BNF . . . 34
11.2 Fill Rules . . . 35
11.2.1 TEMPLATE Object . . . 35
11.2.1.1 DOC_NR Slot . . . 35
11.2.1.2 CONTENT Slot . . . 35
11.2.2 LAUNCH_EVENT Object . . . 36
11.2.2.1 VEHICLE_INFO Slot . . . 36
11.2.2.2 PAYLOAD_INFO Slot . . . 36
11.2.2.3 LAUNCH_DATE slot . . . 36
11.2.2.4 LAUNCH_SITE Slot . . . 37
11.2.2.5 MISSION_TYPE Slot . . . 37
11.2.2.6 MISSION_FUNCTION Slot . . . 37
11.2.2.7 MISSION_STATUS Slot . . . 38
11.2.3 VEHICLE_INFO Object . . . 38
11.2.3.1 VEHICLE Slot . . . 38
11.2.3.2 VEHICLE_TYPE Slot . . . 38

11.2.3.3 VEHICLE_OWNER Slot . . . 39
11.2.3.4 VEHICLE_MAN Slot . . . 39
11.2.4 PAYLOAD_INFO Object . . . 39
11.2.4.1 PAYLOAD Slot . . . 39
11.2.4.2 PAYLOAD_IDENTIFIER Slot . . . 40
11.2.4.3 PAYLOAD_TYPE Slot . . . 40
11.2.4.4 PAYLOAD_FUNC Slot . . . 40
11.2.4.5 PAYLOAD_OWNER Slot . . . 40
11.2.4.6 PAYLOAD_MAN Slot . . . 41
11.2.4.7 PAYLOAD_ORIGIN Slot . . . 41
11.2.4.8 PAYLOAD_TARGET Slot . . . 41
11.2.4.9 PAYLOAD_RECIP Slot . . . 41
11.2.5 TIME Object . . . 42

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, a 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 "one senior Navy officer, who asked to remain anonymous" (alternative fill is "one senior Navy officer")

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 General Template Relation Object

DEFINITION: Relationship between two entities.

MINIMUM INSTANTIATION CONDITIONS: Objects are Template Elements. See MUC-7 Template Element Task Definition. Objects are in the defined relationship.

SPECIAL USAGE NOTES:

1. Existence dependencies. Entities may be created or cease to exist during the time covered in the article. These changes may have an effect on the instantiation of a relationship.

2. Time dependencies. Only the current relationship is to be reported. Current time can be either the time of the article or the time of the story being told within the article depending on the style.

3. Implication by form. A relationship can be inferred from the format of the article.

4. Inference versus world knowledge. The extracted information must be locatable or linked to something in the text. The use of world knowledge in inferencing from information contained in the text is expected during information extraction, but information outside of the text should not be entered in the fills. This distinction is tenuous, but it is reasonable and necessary. The evaluation is not a measure of the system's amount of world knowledge, but a measure of its ability to extract specified material from the text.

5.Valencies. Some entities may be in more than one relationship. Each object represents one relationship.

6. Granularity. Some entities may be proper parts of larger entities. The usage notes will specify which level on this hierarchy should be chosen to represent the relationship. The annotators may decide to allow leniency by putting alternatives or optionals here to avoid the linchpin effect.

10.2.2 PRODUCT_OF Object

DEFINITION: Relationship between an artifact and an organization entity. The artifact is a product of the organization. An artifact is a product of an organization if that organization created the product for sale.

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.

SPECIAL USAGE NOTES:

1. Existence dependencies. Artifacts that are produced or destroyed during the time of the article can be

reported as products of the organization mentioned as their maker.

2. Time dependencies. Only the current maker of the product is to be reported.

3. Implication by form. A modifier of an artifact name can be reported in the PRODUCT_OF relation. For example, if the only time an airplane's manufacturer is mentioned is inside the artifact name, it can be reported as follows:

```
   ``Here comes my ride!'' Ross shouted as the McDonnell Douglas MD
   Explorer came into sight.

<ENTITY-9601290937-6> :=
    ENT_NAME: "McDonnell Douglas"
    ENT_TYPE: ORGANIZATION
    ENT_CATEGORY: ORG_CO

<ENTITY-9601290937-18> :=
    ENT_NAME: "McDonnell Douglas MD Explorer"
    ENT_TYPE: ARTIFACT
    ENT_DESCRIPTOR: "SUPERCHOPPER"
   / "my ride"
   / "CHOPPER"
    ENT_CATEGORY: ART_AIR

<PRODUCT_OF-9601290937-1> :=
    ARTIFACT: <ENTITY-9601290937-18>
    ORGANIZATION: <ENTITY-9601290937-6>
```

4. Inference versus world knowledge. For example, if GM and Chevrolet are mentioned in an article and a Chevrolet is an artifact, the article must imply that Chevrolet is part of GM for GM to be allowed as the maker. (See 6. below.)

5.Valencies. Some organizations may appear as the maker of more than one artifact. A separate template relation should be instantiated for each artifact.

6. Granularity. If an artifact is made by part of an organization and the larger organization is reported, the largest organization reported must appear as the maker and the part of the organization that made the artifact may be reported, too, but it is not necessary. For example, if both GM and Chevrolet are mentioned in an article with Chevrolet as a division of GM and a Chevrolet is an artifact, the artifact **must** be reported as a product of GM. The key will also allow the system to report the Chevrolet as a product of Chevrolet (a division of GM), but it is not penalized if it does not report it.

10.2.3 EMPLOYEE_OF Object

DEFINITION: A relationship between a person and an organization entity. The person is an employee of the organization, that is, the person works for the organization in return for financial compensation.

MINIMUM INSTANTIATION CONDITIONS: Both the ORGANIZATION and PERSON entities must appear in the text and be filled as template elements. The person must work for the organization in return for financial compensation. The relation points to these objects.

SPECIAL USAGE NOTES:

1. Existence dependencies. If a person dies while in the employment of an organization, the EMPLOYEE_OF relationship is to be reported. For example, a stewardess of a TWA plane that crashes who is killed in the crash is still reportable as an employee of the airline.

2. Time dependencies. Only the current employment relationship is to be reported. Current time can be either the time of the article or the time of the story being told within the article depending on the style. If someone has changed companies or retired prior to the time of the article or story, the EMPLOYEE_OF relationship is

not reportable. For example, Edward L. Beach is not an employee of the Navy at the time reported on in the article of 2 January 1996.

```
<ENTITY-9601020516-17> :=
    ENT_NAME: "Edward L. Beach"
    ENT_TYPE: PERSON
    ENT_DESCRIPTOR: "a retired Navy captain and author of a 1994 book
   defending Kimmel and Short"
   / "a retired Navy captain"
   / "author of a 1994 book defending Kimmel and Short"
    ENT_CATEGORY: PER_CIV
 / PER_MIL
```

Time dependencies can be unclear. For example, both positions seem to be current in

```
"Mary Smith, [president of IBM] was promoted to [CEO of AT&T]"
```

The MUC-7 Coreference Task Definition discusses time dependencies, but there is no satisfactory treatment in any of the MUC-7 task definitions including this one.

3. Implication by form. A relationship can be inferred from the format of the article. In the keys, the writer of an article is linked to the newspaper even though there is no statement that the journalist is an employee of the newspaper. For example, the following preamble is enough to make the EMPLOYEE_OF relationship for the author valid.

```
<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>

<EMPLOYEE_OF-9602040136-2> :=
    PERSON: <ENTITY-9602040136-8>
    ORGANIZATION: <ENTITY-9602040136-3>
<ENTITY-9602040136-8> :=
    ENT_NAME: "JOHN O'NEIL"
    ENT_TYPE: PERSON
    ENT_CATEGORY: PER_CIV
<ENTITY-9602040136-3> :=
    ENT_NAME: "N.Y. Times News Service"
    ENT_TYPE: ORGANIZATION
    ENT_CATEGORY: ORG_CO
```

4. Inference versus world knowledge. The extracted information must be locatable or linked to something in the text. The use of world knowledge in inferencing from information contained in the text is expected during information extraction, but information outside of the text should not be entered in the fills. So just because the system may know that Michael Jordan plays for the Chicago Bulls, if there is nothing in the text that implies this relationship, then the relationship is not reportable. However, if a human reader can infer from the context that a player is on a professional team, the relationship must be reported. For example, Aldred should be reported as an employee of the Tigers because of the following excerpt:

```
Gross fared better than Aldred. Tigers manager Buddy Bell announced after
   the game Aldred was being sent to the bullpen.
```

5. Valencies. Some entities may be in more than one relationship. An organization has many employees and a consultant can be an employee of several companies at the same time. Each relationship should be reported as a separate object due to complications of cross reference that would arise if the valencies in the BNF were larger than one fill per slot for the relations.

6. Granularity. Some organizations may be proper parts of a larger organization. An employee can be reported in relations to both the smaller and larger organizations, but, together with the rules for optionality, the EMPLOYEE_OF relation with the smaller organization will be optional. For example, Andrew Blake is related to the Boston Globe in the preamble and it is stated in a sentence at the end of the article that he is on the Globe Staff. He is technically an employee of both organizations. However, the Globe Staff is an optional organization because it is a part of the Boston Globe, so his relationship to the Globe Staff will essentially be optional.

```
<ENTITY-9601120741-11> :=
    ENT_NAME: "the Globe Staff"
    ENT_TYPE: ORGANIZATION
    ENT_CATEGORY: ORG_CO
    OBJ_STATUS: OPTIONAL
    COMMENT: "optional because part of a larger organization"
```

7 Affiliations. A person's general affiliation with an organization is not enough to qualify as an EMPLOYEE_OF relation.

a. club membership. For example, a member of the Boy Scouts is not an employee, but a paid person on the staff of the Boy Scouts of America is an employee.

b. committee membership. If a committee member is paid for their work, it is usually by a different organization than the committee. For example, "a spokesman for the House Ways and Means Committee" is an employee of the House if the House appears in the text separately and not an employee of the committee under any conditions. Technically the Democratic National Committee probably does have employees, but for our purposes, wes assume that committees never pay their members for serving on the committee.

c. team membership. Team members are only considered employees of the team if they are paid by the team. So members of teams in the NBA, NFL, and NHL are in an EMPLOYEE_OF relationship with their teams, but the members of the U.S. men's gymnastics team are not.

8 Ownership. An owner of an organization is generally not considered an employee. It is true that an owner does work and does potentially receive financial benefits, but there is no direct relationship between those benefits and the owner's efforts. However, if the owner also has a paid position in the company, for example, CFO, then they are an employee. So "a founder and CEO of Valujet" would be considered an employee.

10.2.4 LOCATION_OF Object

DEFINITION: A relationship between a LOCATION and an ORGANIZATION. The location of an organization is a specification of its position or limits.

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

SPECIAL USAGE NOTES:

1. Existence dependencies. Entities may be created or cease to exist during the time covered in the article. These changes may have an effect on the instantiation of a relationship.

2. Time dependencies. Only the current relationship is to be reported. Current time can be either the time of the article or the time of the story being told within the article depending on the style.

3. Implication by form. A relationship can be inferred from the format of the article. Fore example, the following beginning paragraph causes an optional LOCATION_OF relation to be instantiated for Navy.

```
   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.

<ENTITY-9602040136-1> :=
    ENT_NAME: "NAVY"
    ENT_TYPE: ORGANIZATION
    ENT_CATEGORY: ORG_GOVT
<LOCATION-9602040136-1> :=
    LOCALE: "WASHINGTON"
    LOCALE_TYPE: CITY
    COUNTRY: "United States"
<LOCATION_OF-9602040136-1> :=
    LOCATION: <LOCATION-9602040136-1>
    ORGANIZATION: <ENTITY-9602040136-1>
    OBJ_STATUS: OPTIONAL
    COMMENT: "implied"
```

However, it cannot be assumed that all organizations in the text are located in Washington. The opening action does seem to take place in the location where the report originates.

4. Inference versus world knowledge. The extracted information must be locatable or linked to something in the text. The use of world knowledge in inferencing from information contained in the text is expected during information extraction, but information outside of the text should not be entered in the fills. For example, the Pentagon should not always be instantiated as being in Washington just because its location is known to the system.

5.Valencies. Some organizations may have multiple locations, such as the Navy or a large or multi-national company. Each of these LOCATION_OF relations is to be instantiated separately.

6. Granularity. Some locations may be proper parts of larger reported locations. In these cases, the most specific location for an organization is the only one that should be instantiated.

10.3 Scoring Note

If a Template Relation in the key points to any optional Template Element, that relationship is automatically made optional by the scorer to limit annotator error.

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 air vehicle launch reports and updates.The template captures information about the launch vehicle, the payload of that vehicle, the date and site of the launch, and information about the mission type, function and status. The event object captures vehicle information, payload information, the date and location of the launch, and mission information. The relational and low-level objects capture vehicle information, including vehicle type, owner and manufacture information, payload information, including information about payload type and function, owner, manufacturer, origin, target, and recipient. A relevant article refers to a vehicle launch that is scheduled, in progress or has actually occurred and must minimally identify the payload, the date of the launch, whether the launch is civilian or military, the function of the mission and its status. The date of the launch should be specified directly or can be inferred from information in the article.

11.1 Template BNF

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

<LAUNCH_EVENT> :=
      VEHICLE_INFO:<VEHICLE_INFO>
      PAYLOAD_INFO:<PAYLOAD_INFO> +
      LAUNCH_DATE:<TIME> ^
      LAUNCH_SITE:<LOCATION>
      MISSION_TYPE:{ MILITARY, CIVILIAN } ^
      MISSION_FUNCTION:{ TEST, DEPLOY, RETRIEVE } +
      MISSION_STATUS:{ SUCCEEDED, FAILED, IN_PROGRESS, SCHEDULED } ^
      COMMENT:    "COMMENT"

<VEHICLE_INFO> :=
      VEHICLE:    <ENTITY>
      VEHICLE_TYPE:{ SHUTTLE, ROCKET, GROUND_BASED, AIRPLANE, SHIP,
SUBMARINE } ^
      VEHICLE_OWNER:<ENTITY>
      VEHICLE_MAN:<ENTITY>
      COMMENT:    "COMMENT"

<PAYLOAD_INFO> :=
      PAYLOAD:    <ENTITY>
      PAYLOAD_IDENTIFIER: "TEXT"
      PAYLOAD_TYPE:{ SATELLITE, MISSILE, SPACE_PROBE, MATERIAL, PERSONNEL,
OTHER } ^
      PAYLOAD_FUNC:{ COMMS, INTELL, WEATHER, TV, DESTRUCTION, RESEARCH, UNK
} ^
      PAYLOAD_OWNER:<ENTITY>
      PAYLOAD_MAN:<ENTITY>
      PAYLOAD_ORIGIN:<LOCATION>
      PAYLOAD_TARGET:<LOCATION>
      PAYLOAD_RECIP:<ENTITY> *
      COMMENT:    "COMMENT"

<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:    "TEXT"

<LOCATION> :=
      LOCALE:     "LOCALE" +
      LOCALE_TYPE:{ CITY, PROVINCE, COUNTRY, REGION, WATER, AIRPORT, UNK } ^
      COUNTRY:    "'yourpath/countries_regions'"
      OBJ_STATUS: { OPTIONAL }
      COMMENT:    "TEXT"

<TIME> :=
      START:      [TEXT]
      END:        [TEXT]
      DESCRIPTOR: "DESCRIPTOR" ^
      OBJ_STATUS: { OPTIONAL }
      COMMENT:    "COMMENT"

%%

DATE-EXP ::[[01-31]][[01-12]][[1000-9999]YYYY]
                  | {FA, WI, SP, SU, 1Q, 2Q, 3Q, 4Q, 1F, 2F, 3F, 4F, FY}
                    [[1000-9999]YYYY]
                  | {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 mission launch event. See *Information Extraction Task Definition* for general information.

MINIMUM INSTANTIATION CONDITIONS: The article must identify, imply, or refer to a vehicle launch, must identify the payload of the vehicle, the launch date, and the mission type, function and status. To be relevant, it must be the case that either the launch event has actually occurred, and that a date can be specified or inferred for it, or that a launch is being scheduled and that a date for the scheduled launch is specified directly or can be inferred from the article or that a mission resulting from a launch event is in progress. The actual launch vehicle itself need not be identified nor information about it provided. More than one launch 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 launches are limited to those that have actually occurred or those that are scheduled, with a

specific date (at least year) either given or inferable for the launch. Hypothetical discussions of a launch ("what if...") are not relevant.
2. More than one launch event may be mentioned in an article. For a second launch event to be relevant, it must be related in mission, payload, or vehicle information to the first relevant launch event.
3. There must be one fully instantiated LAUNCH_EVENT object for the article to be relevant. Additional LAUNCH_EVENT objects must meet the same minimum instantiation conditions.

11.2.2 LAUNCH_EVENT Object.

DEFINITION: The vehicle launch event.

MINIMUM INSTANTIATION CONDITIONS: See section 11.2.1. above.

SPECIAL USAGE NOTES:
1. Instantiate separate LAUNCH_EVENTS for each launch, subject to instantiation conditions.
2. Instantiate separate LAUNCH_EVENT objects for each launch a vehicle may participate in. A vehicle may be involved in more than one LAUNCH_EVENT (e.g. several launches of the same shuttle vehicle, like the Endeavor).
3. In some articles, mention of a launch event may not meet minimum instantiation conditions, e.g. an article discussing the deployment of Sputnik might not provide date information since that information is part of the general knowledge of the reader, or an article discussing satellite sales might mention several future satellite launches, but not provide any date information since that is not the focus of the article. For the sake of consistency, there will be no LAUNCH_EVENT object generated for such launches.

11.2.2.1 VEHICLE_INFO Slot

DEFINITION: Pointer to object carrying information about vehicle type, vehicle, owner, and manufacturer of the vehicle.

MINIMUM INSTANTIATION CONDITIONS: This slot may be empty.

SPECIAL USAGE NOTES:
1. If the payload is a missile and no other launch vehicle is specified, then instantiate a VEHICLE_INFO with the missile as the VEHICLE and VEHICLE_TYPE as ROCKET.

11.2.2.2 PAYLOAD_INFO Slot

DEFINITION: Pointer to object carrying information about the payload, payload type, function, owner, manufacturer, origin, target, and recipient involved in the launch.

MINIMUM INSTANTIATION CONDITIONS: The text must supply a fill for at least one of the following slots: PAYLOAD or PAYLOAD_IDENTIFIER.

SPECIAL USAGE NOTES:
1. Instantiate one PAYLOAD_INFO for each payload in a given LAUNCH_EVENT, subject to instantiation conditions. Several PAYLOAD_INFOs may be instantiated for a given LAUNCH_EVENT, since a launch vehicle may carry several payloads, cf. space shuttle articles.
2. If the launch vehicle carries another vehicle as payload, then any cargo carried on board the non-launch vehicle should not be considered payload and a PAYLOAD_INFO slot will not be instantiated. For example, if the space shuttle carries and deploys a satellite and that satellite carries a tether or an inflatable antenna, which will itself be deployed, do not instantiate a PAYLOAD_INFO slot for the tether or antenna.
3. PAYLOAD_INFO will not be instantiated for payload that is not retrieved to, deployed from, or tested outside of the launch vehicle. For example, PAYLOAD_INFO will not be instantiated for test equipment resting inside, and remaining inside, the shuttle's cargo bay.

11.2.2.3 LAUNCH_DATE slot

DEFINITION: Pointer to an object indicating the date of the vehicle launch.

MINIMUM INSTANTIATION CONDITIONS: This slot must always be filled.

SPECIAL USAGE NOTES: See section 11.2.5 and time-guidelines-supplement-07aug97.
1. If two mentions of dates for the launch 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 launch, choose “January 17”.
2. For scheduled launches, if the launch is scheduled for later in the current day and no specific mention is made in the article of the launch date, the default LAUNCH_DATE will be the date of the article. Otherwise the future date of the scheduled launch must be specified in the article or must be inferable from information in the article (see notes 3-5 below).
3. For articles about the space shuttle, if both the date of the shuttle’s launch and the dates of the payload’s deployment(s) are given, use the date of the shuttle launch as the LAUNCH_DATE.
4. For articles about the space shuttle, if no LAUNCH_DATE for the shuttle is given or inferable, but deployment dates for individual payloads are given (this usually happens when the missions are in progress), use the earliest deployment date as the LAUNCH_DATE.
5. Similarly, for articles not about the shuttle, if there is no mention of a LAUNCH_DATE, but a deployment date is given (i.e. vehicle information is not provided, but payload information is), the LAUNCH_DATE is the same as the earliest deployment date. For example, this would mean that a. rocket launch and deployment of a satellite from that rocket occur on the same day.

11.2.2.4 LAUNCH_SITE Slot

DEFINITION: Pointer to an object indicating the location of the launch

MINIMUM INSTANTIATION CONDITIONS: This slot may be empty.

SPECIAL USAGE NOTES:
1. The most specific location information should be used. For example, in “the rocket was launched from Xichang in the Sichuan Province in southwestern China”, then LAUNCH_SITE will be a pointer to the LOCATION object for “Xichang.”

11.2.2.5 MISSION_TYPE Slot

DEFINITION: A general categorization of the type of mission being performed.

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 launch of a military payload.
CIVILIAN -- A launch of a civilian payload.
2. If no specific mention of a military mission is given, the default is CIVILIAN. For example, the deployment of Sputnik would be of MISSION_TYPE CIVILIAN if no mention of military mission is provided in the article.
3. The selection of the MISSION_TYPE is dependent on the payload being launched. For example, if a military intelligence satellite is being launched from a civilian vehicle (e.g. the space shuttle), the MISSION_TYPE is MILITARY.
4. If more than one payload is being launched from the same vehicle (e.g. a space shuttle), then military payload will take precedence and determine the MISSION_TYPE. For example, if both a commercial communications satellite and a military intelligence satellite are being launched, the MISSION_TYPE will be MILITARY, even though one or more of the payloads is civilian.
5. Civilian government organizations, such as NASA, are considered CIVILIAN.

11.2.2.6 MISSION_FUNCTION Slot

DEFINITION: A categorization of the mission of the launch with respect to the payload. Multiple MISSION_FUNCTIONS may be specified.

MINIMUM INSTANTIATION CONDITIONS: This slot must be filled.

SPECIAL USAGE NOTES:
1. The categories that are to be used for this slot are defined as follows:
TEST -- The mission is to TEST the payload.
DEPLOY -- The mission is to DEPLOY the payload.
RETRIEVE.-- The mission is to RETRIEVE the payload
2. More than one MISSION_FUNCTION category may be selected. For example, if on a single mission, from a single launch vehicle, one satellite is being deployed and a different satellite is being retrieved, both

DEPLOY and RETRIEVE will be selected.
3. If a payload is deployed, tested, and then retrieved on the same mission, then the MISSION_FUNCTION associated with that payload will be TEST.
4. The default value of MISSION_FUNCTION is DEPLOY, if no other MISSION_FUNCTION is specified or inferred from information the article.

11.2.2.7 MISSION_STATUS Slot

DEFINITION: A categorization of the status of the mission with respect to its success. Only one may be selected.

MINIMUM INSTANTIATION CONDITIONS: This slot must be filled.

SPECIAL USAGE NOTES:
1. The categories that are to be used for this slot are defined as follows:
SUCCEEDED -- The mission function has been fulfilled.
FAILED -- The mission function has not been fulfilled.
IN_PROGRESS -- The mission function has not yet been completed.
SCHEDULED -- The mission has not yet occurred and has been scheduled.
2. All mission functions must be successful for SUCCEEDED to be specified. For example, if a mission includes the deployment of several payloads, then all payloads must have been successfully deployed for the MISSION_STATUS to be SUCCEEDED.
3. For MISSION_STATUS to be FAILED, at least one of the mission functions associated with its payloads must have failed. For example, if a shuttle mission involves two satellite deployments and one has failed while the other has succeeded, then the category selected is FAILED. In another example, if the mission involves testing a satellite by first deploying it and then retrieving it, if one of the actions has failed, e.g. the retrieval failed, then the mission is not successful and the fill is FAILED.
3. For MISSION_STATUS to be IN_PROGRESS, the launch must have already occurred, but a MISSION_FUNCTION related to a particular payload is not complete, e.g. the satellite being tested has been deployed, but not yet retrieved. This category is only appropriate for shuttle missions and space station missions.
4. For MISSION_STATUS to be SCHEDULED, a future date for the launch must be specified or inferred.
5. If the mission is only partially successful at its conclusion, e.g. “completed 70% of its objectives”, then MISSION_STATUS is FAILED.
6. If the launch involves missiles, assume the MISSION_STATUS is SUCCEEDED, unless otherwise noted, i.e. the missile has hit its target.
7. If a mission has not yet been completed, and one portion of the mission has not been successful, mark the MISSION_STATUS as FAILED.

11.2.3 VEHICLE_INFO Object

DEFINITION: This object relates information about the launch vehicle.

MINIMUM INSTANTIATION CONDITIONS: See section 11.2.21.

11.2.3.1 VEHICLE Slot

DEFINITION: Pointer to the object that identifies and names the vehicle involved in the launch event.

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

SPECIAL USAGE NOTES:
1. Text must directly identify the vehicle model or series or else the vehicle must be identified by a substantial descriptor, as defined in the *Information Extraction Task Definition*. For example, a vehicle model or series would be things like J-802 or Long March or Saturn.

11.2.3.2 VEHICLE_TYPE Slot

DEFINITION: A categorization of the type of launch vehicle. Only one may be selected.

MINIMUM INSTANTIATION CONDITIONS: This slot must be filled.

SPECIAL USAGE NOTES:
1. The categories that are to be used for this slot are defined as follows:
SHUTTLE -- The space shuttle.
ROCKET -- A rocket.
GROUND_BASED -- Vehicle is launched from the ground. Used for some kinds of missiles.
AIRPLANE -- A commercial, private, or military aircraft.
SHIP -- A vehicle that rides on the surface of water.
SUBMARINE -- An underwater vehicle.
2. If the article is about the launch of a space shuttle and its payload and no specific mention of the rocket used to launch the space shuttle is made, consider the shuttle to be the launch vehicle. If mention is made of a particular rocket carrying a shuttle, then specify the launch vehicle as the rocket and not the shuttle.
For example, in "NASA is proceeding with its plans to launch the shuttle Endeavor on a mission to retrieve a Japanese science satellite," the VEHICLE_TYPE is SHUTTLE. In "Hyflex, Japan's space shuttle prototype, will shoot up through the atmosphere on its Japanese-made J1 booster rocket," the VEHICLE_TYPE is ROCKET [and the VEHICLE is the J1].

11.2.3.3 VEHICLE_OWNER Slot

DEFINITION: Pointer to the object that identifies the owner of the vehicle. 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. VEHICLE_OWNER will be filled by a pointer to an ENTITY of type ORGANIZATION or PERSON.
2. When VEHICLE_OWNER is filled by a pointer to an ENTITY object of type ORGANIZATION. The organization can be a private company, a branch of the military (such as the "Navy"), or a civilian government organization (such as "NASA").
3. VEHICLE_OWNER may be instantiated by a launch service operator or a launch agent, e.g. Great Wall.

11.2.3.4 VEHICLE_MAN Slot

DEFINITION: Identification of the company that manufactured the launch vehicle.

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 vehicle is military, the manufacturer must be identified independently of the military in order for this slot to be filled.

11.2.4 PAYLOAD_INFO Object

DEFINITION: This object relates information about the payload for the launch.

MINIMUM INSTANTIATION CONDITIONS: See section 11.2.2.2 above.

11.2.4.1 PAYLOAD Slot

DEFINITION: Pointer to the object that identifies and names the payload being carried by the launch vehicle.

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

SPECIAL USAGE NOTES:
1. Text must directly identify the name of the payload (if the payload is human) or the payload model or series, or else a substantial descriptor must be provided.
2. If a launch vehicle itself is being tested, and the vehicle is either not carrying a specific payload or is carrying no payload, then the PAYLOAD is the same as the VEHICLE, i.e. the vehicle itself is the payload during a test of the launch vehicle. For example, if an Ariane 5 rocket is launched as a test vehicle, then both

the PAYLOAD and the VEHICLE will be instantiated with Ariane 5.
3. Personnel are not considered PAYLOAD if they are deploying, retrieving, or testing payloads, like satellites. So, if a person conducts a space walk while deploying a satellite, he/she is not considered the PAYLOAD.

11.2.4.2 PAYLOAD_IDENTIFIER Slot

DEFINITION: Noun phrase identifying or referring to a payload that is not of type vehicle or person. This slot is not permitted to have more than one value.

MINIMUM INSTANTIATION CONDITIONS: Text must provide a string that identifies a payload and that does not fit the definition of PAYLOAD slot.

SPECIAL USAGE NOTES:
1. This is intended to capture information on cargo that is being deployed, retrieved or tested outside the launch vehicle. A payload that is being tested or merely rests within the launch vehicle, e.g. within the Shuttle cargo bay, will not qualify as a PAYLOAD_IDENTIFIER.

11.2.4.3 PAYLOAD_TYPE Slot

DEFINITION: A categorization of the type of payload being carried by the launch vehicle. One must be selected.

MINIMUM INSTANTIATION CONDITIONS: This slot must be filled.

SPECIAL USAGE NOTES:
1. The categories that are to be used for this slot are defined as follows:
SATELLITE -- A satellite.
MISSILE -- A missile, land-, sea-, or air-launched.
SPACE_PROBE -- A space probe.
MATERIAL -- Supplies or equipment.
PERSONNEL -- A person.
OTHER -- Some payload other than those specified above.
1. Only one item may be selected from PAYLOAD_TYPE.
2. PAYLOAD_TYPE will be PERSONNEL if a specific person is being deployed to or retrieved from another vehicle or location (e.g. transferred to space station MIR, landing on the moon) or if the person will participate as a subject in an experiment. If no mention is made of any of these kinds activities, then personnel will not be considered as PAYLOAD and a PAYLOAD_TYPE will not be instantiated.

11.2.4.4 PAYLOAD_FUNC Slot

DEFINITION: A categorization of the function or use of payload being carried by the launch vehicle. One or more must be selected.

MINIMUM INSTANTIATION CONDITIONS: This slot must be filled.

SPECIAL USAGE NOTES:
1. The categories that are to be used for this slot are defined as follows:
COMMS -- payload is for transmitting or receiving communications, such as cellular telephone traffic, e.g. communications satellite
INTELL -- payload receives or transmits intelligence information, e.g. intelligence satellite
WEATHER -- payload receives or transmits weather information, e.g. weather satellite
TV -- payload transmits or receives TV signals, e.g. digital TV satellite
DESTRUCTION -- payload is used for destructive purposes, e.g. missiles.
RESEARCH -- payload is being used to conduct research, e.g. satellite collecting info in cosmic rays or a space probe being sent off for journey to Mars.
UNK -- function of payload is unknown from information in the article.

11.2.4.5 PAYLOAD_OWNER Slot

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

MINIMUM INSTANTIATION CONDITIONS: This slot may be empty.

SPECIAL USAGE NOTES:
1. For CIVILIAN payload, the PAYLOAD_OWNER will be filled by a pointer to an ENTITY of type ORGANIZATION or PERSON. Note that civilian government organizations like NASA are considered CIVILIAN for this task.
2. For MILITARY payload, PAYLOAD_OWNER 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").
3. If the payload is of PAYLOAD_TYPE PERSONNEL, the PAYLOAD_OWNER slot will be instantiated with the employer of that person. For example, if Shannon Lucid is the payload, then the PAYLOAD_OWNER will be NASA.

11.2.4.6 PAYLOAD_MAN Slot

DEFINITION: Identification of the company that manufactured the payload.

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 payload is military, the manufacturer must be identified independently of the military in order for this slot to be filled.

11.2.4.7 PAYLOAD_ORIGIN Slot

DEFINITION: Pointer to an object indicating the location of payload's manufacturer or owner.

MINIMUM INSTANTIATION CONDITIONS: This slot may be empty.

SPECIAL USAGE NOTES:
1. The most specific location information should be used. For example, if the article contains information about a satellite manufactured in Italy by Manufa and also contains information that Manufa is in Milan, then use Milan as the PAYLOAD_ORIGIN.
2. If no specific manufacturer or owner location information is available, but location information can be extracted, e.g. in "Italian satellite", fill the PAYLOAD_ORIGIN with the location, e.g. "Italy".
3. If location information about both the manufacturer and owner are available, use the location of the owner.

11.2.4.8 PAYLOAD_TARGET Slot

DEFINITION: Identification of the target location or destination for the payload.

MINIMUM INSTANTIATION CONDITIONS: This slot may be empty.

SPECIAL USAGE NOTES:
1. This slot is filled by a pointer to a LOCATION object.
2. If the PAYLOAD_TYPE is a MISSILE, then the PAYLOAD_TARGET is an earth-based LOCATION, such as a city, nation, or region. If the PAYLOAD_TYPE is a SPACE_PROBE, then the PAYLOAD_TARGET is space-based, such as a planet, or asteroid.
3. If the payload has several destinations scheduled (e.g. a space probe might have several intermediary destinations and then a final destination), then only the final destination should be extracted.

11.2.4.9 PAYLOAD_RECIP Slot

DEFINITION: Identification of the recipient(s) of the services of the payload. This slot may have more than one fill.

MINIMUM INSTANTIATION CONDITIONS: This slot may be empty.

SPECIAL USAGE NOTES:
1. PAYLOAD_RECIP is a pointer to an ENTITY object. There may be pointers to more than one ENTITY.
2. The name or description of the recipient must be given.

3. The recipient can be a pointer to an ORGANIZATION. For example, in "News Corp. leased space on the Intelsat satellite," the PAYLOAD_RECIP is "News Corp." If several organizations are specified as leasing space, all will be identified in the PAYLOAD_RECIP slot. However, if the organizations have joined in a consortium and that consortium is named, then the consortium name will be extracted and the names of the individual consortium members will not be identified.
4. The recipient can be a pointer to an ARTIFACT. For example, in "one missile struck the frigate Stark", the ENTITY filled by "the frigate Stark" is the PAYLOAD_RECIP. Another example is found in "the astronaut was transferred to the space station MIR," Here MIR would be the PAYLOAD_RECIP.
5. The recipient can be a pointer to a PERSON. For example, Timothy Leary and Gene Rodenberry are PAYLOAD_RECIPs as they receive the services of the satellite that is launching their ashes into outer space. Similarly, an individual may be a PAYLOAD_RECIP is he or she is the object of a missile launch.
6. The payload owner will not be instantiated as a PAYLOAD_RECIP unless there is some other recipient as well. In most cases, payload owner will not be instantiated as a payload recipient.

11.2.5 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:
These special usage notes supersede remarks in the time-guidelines-supplement-07aug97.
1. Interpretation of the template element may require calculation from the reference date of the article. For example, "in the next 2 days" means that the START value is today, and the END value is today+2.
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 same day, e.g. Saturday morning vs. Saturday evening, different TIME objects will be instantiated. If only a day of the week is specified for both events, e.g. Saturday, then the one TIME object 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.
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

9. For time expressions where specific start and end points are not given, for example "in April" or "during June," only the month and year need be given.

TIME: START: 041997
END: 041997

10. DESCRIPTOR will not include a preposition unless that preposition changes the semantics of the phrase. This is consistent with NE markup requirements. For example, “since 1995” would be a DESCRIPTOR because “1995” and “since 1995” have two different meanings. Similarly “between Thursday and Saturday” would be a DESCRIPTOR. For the prepositional phrase “on Tuesday,” “Tuesday” would be the DESCRIPTOR.

11. Temporal modifiers (cf. “A Revised Template Description for Time (v3)”) have been removed from the Template BNF in section 11.1 above. For TIME expressions that contain “early, mid, late”, e.g. early May, mid-April, late 1997, appropriate dates will be calculated. Divide the time period (month, year) into three equal parts, the first third = early, the second third = mid, and the last third = late. If there is an extra unit (month of 31 days), tack the remainder onto the “late” range. For example, for months the following would be appropriate:

TIME: START: 01051997
END: 10051997
DESCRIPTOR: “early May 1997”

TIME: START: 11051997
END: 20051997
DESCRIPTOR: “mid May 1997”

TIME: START: 21051997
END: 31051997
DESCRIPTOR: “late May 1997”

For years, the following would be appropriate:

TIME: START: 011997
END: 041997
DESCRIPTOR: “early 1997”

TIME: START: 051997
END: 081997
DESCRIPTOR: “mid 1997”

TIME: START: 091997
END: 121997
DESCRIPTOR: “late 1997”

Decades, centuries, etc. should be handled similarly.