"unchanged at 95.05"
`unchanged at 95.05`
[no markup]

"1.5 times"
`1.5 times`
[no markup]

"about one-third of"
`about one-third of`
[no markup]

"Fees 1 3/4."
`Fees 1 3/4.`
[no markup]

"a fixed 106 7/8"
`a fixed 106 7/8`
[no markup]

"priced at 99 1/4"
`priced at 99 1/4`
[no markup]

APPENDIX C. NUMEX: SPECIFIC GUIDELINES

C.1 The entire string expressing the monetary or percentage value is to be tagged.

"20 million New Pesos"
```
<NUMEX TYPE="MONEY">20 million New Pesos</NUMEX>
```

"$42.1 million"
```
<NUMEX TYPE="MONEY">$42.1 million</NUMEX>
```

"million-dollar conferences"
```
<NUMEX TYPE="MONEY">million-dollar</NUMEX> conferences
```

"15 pct"
```
<NUMEX TYPE="PERCENT">15 pct</NUMEX>
```

C.2 Juxtaposed strings expressing monetary values in two different currencies are to be tagged separately.

"#26 million ($43.6 million)"
```
<NUMEX TYPE="MONEY">#26 million</NUMEX> (<NUMEX TYPE="MONEY">$43.6
   million</NUMEX>)
```

C.3 Modifying words that indicate the approximate value of a number or a "relative position" to a number are to be excluded from the NUMEX tag. However, borderline cases surh as the quantifier "several" may be optionally included in the tagged string.

"about 5%"
```
about <NUMEX TYPE="PERCENT">5%</NUMEX>
```

"over $90,000"
```
over <NUMEX TYPE="MONEY">$90,000</NUMEX>
```

"several million dollars"
```
<NUMEX TYPE="MONEY" ALT="million dollars">several million dollars</NUMEX>
```

C.4 Rarely, multiword strings that are to be tagged as NUMEX will contain LOCATION (ENAMEX) substrings that are also to be tagged.

"the equivalent of less than a U.S. penny"
```
the equivalent of less than <NUMEX TYPE="MONEY">a <ENAMEX
   TYPE="LOCATION">U.S.</ENAMEX> penny</NUMEX>
```

"20 million New Taiwan Dollars"
```
<NUMEX TYPE="MONEY">20 million New </ENAMEX TYPE="LOCATION">Taiwan</ENAMEX>
   Dollars</NUMEX>
```

"US$43.6 million"
```
<NUMEX TYPE="MONEY"><ENAMEX TYPE="LOCATION">US</ENAMEX>$43.6
   million</NUMEX>
```

C.5 Numeric expressions that do not use currency terms to indicate money values and that do not use percentage terms to indicate percentages are not to be tagged.

"12 points"
```
12 points
```
[no markup]

"first-half profit"
```
<TIMEX TYPE="DATE">first-half</TIMEX> profit
```

"fiscal 1989's fourth quarter"
```
<TIMEX TYPE="DATE">fiscal 1989's fourth quarter</TIMEX>
```

"4th period" [of a year]
```
<TIMEX TYPE="DATE">4th period</TIMEX>
```

"1975 World Series"
```
<TIMEX TYPE="DATE">1975</TIMEX> World Series
```

B.3 Relative time expressions are not to be tagged, but any absolute times expressed as part of the entire expression are to be tagged.

"July last year"
```
<TIMEX TYPE="DATE">July</TIMEX> last year
```

"this June"
```
this <TIMEX TYPE="DATE">June</TIMEX>
```

"next summer"
```
next <TIMEX TYPE="DATE">summer</TIMEX>
```

"thirty days before the end of the year"
```
thirty days before the end of the year
```
[no markup]

"the end of 1991"
```
the end of <TIMEX TYPE="DATE">1991</TIMEX>
```
[interpretation of "end" is unclear; it could be a day or month, or something else]

B.4 Special days, such as holidays, that are referenced by name, are not required to be tagged; they will be marked optional in the answer key.

"because of the observance of All Saints' Day"
```
because of the observance of <TIMEX TYPE="DATE" STATUS="OPT">All Saints'
   Day</TIMEX>
```

B.5 Rarely, multiword strings that are to be tagged as TIMEX will contain LOCATION (ENAMEX) substrings that are also to be tagged.

"1:30 p.m. Chicago time"
```
<TIMEX TYPE="TIME">1:30 p.m. <ENAMEX TYPE="LOCATION">Chicago</ENAMEX>
   time</TIMEX>
```

APPENDIX B. TIMEX: SPECIFIC GUIDELINES

B.1 Only "absolute" time expressions are to be tagged. To be considered an absolute time expression, the expression must indicate a specific segment of time, as follows:

- An expression of minutes must indicate a particular minute and hour, such as "20 minutes after 10" (not "a few minutes after the hour," "20 minutes after the hour").
- An expression of hours must indicate a particular hour, such as "midnight," "twelve o'clock noon," "noon" (not "mid-day," "morning").
- An expression of days must indicate a particular day, such as "Monday," "10th of October" (not "first day of the month").
- An expression of seasons must indicate a particular season, such as "autumn" (not "next season").
- An expression of financial quarters or halves of the year must indicate which quarter or half, such as "fourth quarter," "first half." Note that there are no proper names, per se, representing these time periods. Nonetheless, these types of time expressions are important in the business domain and are therefore to be tagged.
- An expression of years must indicate a particular year, such as "1995" (not "the current year").
- An expression of decades must indicate a particular decade, such as "1980s" (not "the last 10 years").
- An expression of centuries must indicate a particular century, such as "the 20th century" (not "this century").

B.2 Absolute time expressions are to be tagged as a single item. Subparts are not to be tagged separately, even in the case of possessive constructions.

"twelve o'clock noon"
```
<TIMEX TYPE="TIME">twelve o'clock noon</TIMEX>
```

"5 p.m. EST"
```
<TIMEX TYPE="TIME">5 p.m. EST</TIMEX>
```

"January 1990"
```
<TIMEX TYPE="DATE">January 1990</TIMEX>
```

"fiscal 1989"
```
<TIMEX TYPE="DATE">fiscal 1989</TIMEX>
```

"the autumn report"
```
the <TIMEX TYPE="DATE">autumn</TIMEX> report
```

"an Indian summer of the soul"
```
<TIMEX TYPE="DATE" ALT="summer">an Indian summer</TIMEX> of the soul
```

"third quarter of 1991"
```
<TIMEX TYPE="DATE">third quarter of 1991</TIMEX>
```

"the fourth quarter ended Sept. 30"
```
<TIMEX TYPE="DATE">the fourth quarter ended Sept. 30</TIMEX>
```

"the three months ended Sept. 30" [as paraphrase of "the fourth quarter ended Sept. 30"]
```
<TIMEX TYPE="DATE">the three months ended Sept. 30</TIMEX>
```

"the first half of fiscal 1990"
```
<TIMEX TYPE="DATE">the first half of fiscal 1990</TIMEX>
```

"Alzheimer's"
Alzheimer's
[no markup]

"the Nobel prize"
the Nobel prize
[no markup]

"St. Michael"
St. Michael
[no markup]

A.4 Guidelines That Pertain Only to LOCATION

A.4.1 The phrase "of <place-name>" following an organization name may or may not be part of the organization name proper. The annotation in the answer key will follow these guidelines: (1) If there is a corporate designator, it marks the end of the organization name; (2) if there is no corporate designator, the "of <place-name>" is part of the organization name.

"Hyundai of Korea, Inc."
<ENAMEX TYPE="ORGANIZATION">Hyundai of Korea, Inc.</ENAMEX>

"Hyundai, Inc. of Korea"
<ENAMEX TYPE="ORGANIZATION">Hyundai, Inc.</ENAMEX> of <ENAMEX
TYPE="LOCATION">Korea</ENAMEX>

"McDonald's of Japan"
<ENAMEX TYPE="ORGANIZATION">McDonald's of Japan</ENAMEX>

"University of California in Los Angeles"
<ENAMEX TYPE="ORGANIZATION">University of California</ENAMEX> in <ENAMEX
TYPE="LOCATION">Los Angeles</ENAMEX>
[where the first tagged string functions as organization rather than facility (see A.2, above, re facilities) -- note also that the locative phrase "in Los Angeles" is not included inside the organization tag]

A.4.2 Compound expressions in which place names are separated by a comma are to be tagged as separate instances of LOCATION.

"Kaohsiung, Taiwan"
<ENAMEX TYPE="LOCATION">Kaohsiung</ENAMEX>, <ENAMEX
TYPE="LOCATION">Taiwan</ENAMEX>

"Washington, D.C."
<ENAMEX TYPE="LOCATION">Washington</ENAMEX>, <ENAMEX TYPE="LOCATION">
D.C.</ENAMEX>

A.4.3 Miscellaneous location-related strings that are not to be tagged as LOCATION include street addresses, street names, adjectival forms of location names.

"53140 Gatchell Road"
53140 Gatchell Road
[no markup]

"American exporters"
American exporters
[no markup]

reference to the facility as a structure/place. Some of the tagged strings may be marked optional in the answer key if the annotator thinks that it's highly unlikely (but not inconceivable) that the name represents an organization.

"Finger Lakes Area Hospital Corp."
<ENAMEX TYPE="ORGANIZATION">Finger Lakes Area Hospital Corp.</ENAMEX>

"Four Seasons Hotels"
[markup as ORGANIZATION likely]

"Unification Church"
[markup as ORGANIZATION likely]

"the White House"
[markup as ORGANIZATION, or no markup]

"Trinity Lutheran Church"
[markup as ORGANIZATION, or no markup]

"General Hospital"
[markup as ORGANIZATION, or no markup]

A.3 Guidelines That Pertain Only to PERSON

A.3.1 Titles such as "Mr." and role names such as "President" are *not* considered part of a person name. However, appositives such as "Jr." *are* considered part of a person name.

"Mr. Harry Schearer"
Mr. <ENAMEX TYPE="PERSON">Harry Schearer</ENAMEX>

"Secretary Robert Mosbacher"
Secretary <ENAMEX TYPE="PERSON">Robert Mosbacher</ENAMEX>

"John Doe, Jr."
<ENAMEX TYPE="PERSON">John Doe, Jr.</PERSON>

A.3.2 Family names are to be tagged.

"the Kennedy family"
the <ENAMEX TYPE="PERSON">Kennedy</ENAMEX> family

"the Kennedys"
the <ENAMEX TYPE="PERSON">Kennedys</ENAMEX>
[alternate form of identification of the Kennedy family entity

A.3.3 Miscellaneous types of proper names that are not to be tagged as PERSON include political groups, laws named after people, diseases/prizes named after people, saints.

"The Republicans held a rally."
The Republicans held a rally.
[no markup]

"the Gramm-Rudman amendment"
the Gramm-Rudman amendment
[no markup]

"Big Board" [alias for New York Stock Exchange]
```
<ENAMEX TYPE="ORGANIZATION">Big Board</ORGANIZATION>
```

"Mr. Fix-It" [nickname for candidate for head of the CIA]
```
Mr. <ENAMEX TYPE="PERSON">Fix-It</ENAMEX>
```

"the Big Apple" [nickname for New York City]
```
<ENAMEX TYPE="LOCATION">the Big Apple</ENAMEX>
```

A.1.7 Miscellaneous types of proper names that are *not* to be tagged as ENAMEX include artifacts, other products, and plural names that do not identify a single, unique entity. (For information on the treatment of facilities, see section A.2, below.)

"Macintosh computers"
```
Macintosh computers
```
[no markup]

"Wall Street Journal"
```
Wall Street Journal
```
[no markup]

"the Campbell Soups of the world"
```
the Campbell Soups of the world
```
[no markup]

A.2 Guidelines That Pertain Only to ORGANIZATION

A.2.1 Corporate designators such as "Co." are considered part of an organization name.

"Bridgestone Sports Co."
```
<ENAMEX TYPE="ORGANIZATION">Bridgestone Sports Co.</ENAMEX>
```

A.2.2 Miscellaneous types of proper names that are to be tagged as ORGANIZATION include sports teams, stock exchanges, multinational organizations, political parties, orchestras, unions, governmental bodies at any level of importance.

"the mayor who built Candlestick Park for the Giants"
```
the mayor who built Candlestick Park for the <ENAMEX
   TYPE="ORGANIZATION">Giants</ENAMEX>
```
[a sports team]

"NASDAQ"
```
<ENAMEX TYPE="ORGANIZATION">NASDAQ</ENAMEX>
```
[a stock exchange]

"European Community"
```
<ENAMEX TYPE="ORGANIZATION">European Community</ENAMEX>
```

"Labor Party"
```
<ENAMEX TYPE="ORGANIZATION">Labor Party</ENAMEX>
```

"Machinists union"
```
<ENAMEX TYPE="ORGANIZATION">Machinists</ENAMEX> union
```

A.2.3 Miscellaneous types of proper names referring to facilities (e.g., factories, hotels, universities, airports, hospitals, churches) will be tagged as ORGANIZATION unless (1) the name is clearly the name of the structure/ place and not of an organization, or (2) the name is known to be the name of an organization but is used only in

"Ford Taurus"
Ford Taurus
[no markup, not even for "Ford"]

"Dow Jones Industrial Average"
Dow Jones Industrial Average
[no markup, not even for "Dow Jones"]

"West Texas Intermediate crude"
West Texas Intermediate crude
[no markup, not even for "West Texas" -- this example differs from those in A.1.1, in that "West Texas Intermediate" is a name and can therefore not be decomposed]

A.1.4 Uncapitalized, common-noun designators such as "division" in the phrase "Chrysler division" are *not* considered part of an entity name.

"Chrysler division"
<ENAMEX TYPE="ORGANIZATION">Chrysler</ENAMEX> division

"the Kennedy family"
the <ENAMEX TYPE="PERSON">Kennedy</ENAMEX> family

"the Southwest region"
the <ENAMEX TYPE="LOCATION">Southwest</ENAMEX> region

A.1.5 In a possessive construction, the possessor and possessed ENAMEX substrings should be tagged separately.

"Temple University's Graduate School of Business"
<ENAMEX TYPE="ORGANIZATION">Temple University</ENAMEX>'s <ENAMEX
TYPE="ORGANIZATION">Graduate School of Business</ENAMEX>

"Shearson Lehman Hutton's OTC department"
<ENAMEX TYPE="ORGANIZATION">Shearson Lehman Hutton</ENAMEX>'s <ENAMEX
TYPE="ORGANIZATION">OTC</ENAMEX> department

"California's Silicon Valley"
<ENAMEX TYPE="LOCATION">California</ENAMEX>'s <ENAMEX
TYPE="LOCATION">Silicon Valley</ENAMEX>

"Canada's Parliament"
<ENAMEX TYPE="LOCATION">Canada</ENAMEX>'s <ENAMEX
TYPE="ORGANIZATION">Parliament</ENAMEX>

A.1.6 Aliases for entities are to be tagged. (For further information on what constitutes an alias, see the "Information Extraction Task Definition.")

"IBM" [alias for International Business Machines Corp.]
<ENAMEX TYPE="ORGANIZATION">IBM</ORGANIZATION>

"Big Blue" [alias for International Business Machines Corp.]
<ENAMEX TYPE="ORGANIZATION">Big Blue</ORGANIZATION>

APPENDIX A. ENAMEX: SPECIFIC GUIDELINES

A.1 Guidelines That Pertain to All Three TYPEs

A.1.1 Entity names used as modifiers in complex NPs that are not proper names are to be tagged when it is clear to the annotator from context or the annotator's knowledge of the world that the name is that of an organization or person.

"Bridgestone profits"
<ENAMEX TYPE="ORGANIZATION">Bridgestone</ENAMEX> profits

"the Clinton government"
the <ENAMEX TYPE="PERSON">Clinton</ENAMEX> government

"Treasury bonds and securities"
<ENAMEX TYPE="ORGANIZATION">Treasury</ENAMEX> bonds and securities

"U.S. exporters"
<ENAMEX TYPE="LOCATION">U.S.</ENAMEX> exporters

A.1.2 In addition, entity names modifying person identifiers (where person identifier= title/role and/or name) are to be tagged.

"Mips Vice President John Hime" [Mips is the name of a computer company]
<ENAMEX TYPE="ORGANIZATION">Mips</ENAMEX> Vice President <ENAMEX
TYPE="PERSON">John Hime</ENAMEX>

"Treasury Secretary"
<ENAMEX TYPE="ORGANIZATION">Treasury</ENAMEX> Secretary

"the U.S. Vice President"
the <ENAMEX TYPE="LOCATION">U.S.</ENAMEX> Vice President

A.1.3 In some cases, multiword strings that are proper names will contain entity name substrings; such strings are not decomposable; therefore, the substrings are not to be tagged. (See A.1.2 re special cases involving prenominal modifiers of person identifiers.)

"Arthur Anderson Consulting"
<ENAMEX TYPE="ORGANIZATION">Arthur Anderson Consulting</ENAMEX>
[no markup for "Arthur Anderson" alone]

"Boston Chicken Corp."
<ENAMEX TYPE="ORGANIZATION">Boston Chicken Corp.</ENAMEX>
[no markup for "Boston" alone]

"U.S. Fish and Wildlife Service"
<ENAMEX TYPE="ORGANIZATION">U.S. Fish and Wildlife Service</ENAMEX>
[no markup for "U.S." alone]

"Northern California"
<ENAMEX TYPE="LOCATION">Northern California</ENAMEX>
[no markup for "California" alone]

"West Texas"
<ENAMEX TYPE="LOCATION">West Texas</ENAMEX>
[no markup for "Texas" alone]

The ALT attribute will be used sparingly. A possible TIMEX example is shown below.

"all of 1987"
```
<TIMEX TYPE="DATE" ALT="1987">all of 1987</TIMEX>
```

4 GUIDELINES FOR MARKUP OF EXCEPTIONAL CONSTRUCTIONS

4.1 Expressions Involving Elision

Multi-name expressions containing conjoined modifiers (with elision of the head of one conjunct) should be marked up as separate expressions.

"North and South America"
```
<ENAMEX TYPE="LOCATION">North</ENAMEX> and
   <ENAMEX TYPE="LOCATION">South America</ENAMEX>
```

A similar case involving elision with number expressions:

"10- and 20-dollar bills"
```
<NUMEX TYPE="MONEY">10</NUMEX>- and
   <NUMEX TYPE="MONEY">20-dollar</NUMEX> bills
```

In contrast, there is no elision in the case of single-name expressions containing conjoined modifiers; such expressions should be marked up as a single expression.

"U.S. Fish and Wildlife Service"
```
<ENAMEX TYPE="ORGANIZATION">U.S. Fish and Wildlife Service</ENAMEX>
```

The subparts of range expressions should be marked up as separate expressions.

"175 to 180 million Canadian dollars"
```
<NUMEX TYPE="MONEY">175</NUMEX> to <NUMEX TYPE="MONEY">180 million Canadian
   dollars</NUMEX>
```

"the 1986-87 academic year"
```
the <TIMEX TYPE="DATE">1986</TIMEX>-<TIMEX TYPE="DATE" ALT="87">87
   academic year</TIMEX>
```

4.2 Effects of Tokenization Conventions

The systems must incorporate certain tokenization conventions. These conventions are contained in a separate document titled "Tokenization Rules."

The tokenization conventions for MUC-6 have an impact on the boundaries of the strings to be tagged. For example, the conventions call for treating possessive forms, e.g., "California's," as multiple tokens, unless there is a name such as "McDonald's [burger company]" that is inherently possessive. See the separate documentation titled "Tokenization Rules" for further information and examples.

4.3 Nested Expressions

There are cases where LOCATION (ENAMEX) expressions are to be tagged when they occur within TIMEX and NUMEX expressions. See Appendices B and C for a description of cases where markup of nested expressions may occur. Entity names that appear within ENAMEX tags are *not* to be tagged.

```
<ELEMENT-NAME ATTR-NAME="ATTR-VALUE" ...>text-string</ELEMENT-NAME>
```

Example:

```
<ENAMEX TYPE="ORGANIZATION">Taga Co.</ENAMEX>
```

The markup is defined in SGML Document Type Descriptions (DTDs), written for MUC-6 use by personnel at MITRE and maintained by personnel at NRaD. The DTDs enable annotators and system developers to use SGML validation tools to check the correctness of the SGML-tagged texts produced by the annotator or the system. The validation tools are available to MUC-6 participants in the file called muc6-sgml-tools. Annotators are using a software tool provided for MUC-6 by SRA Corporation to assist in generating the answer keys to be used for system training and testing.

2.2 Named Entities (ENAMEX tag element)

This subtask is limited to proper names, acronyms, and perhaps miscellaneous other unique identifiers, which are categorized via the TYPE attribute as follows:

ORGANIZATION: named corporate, governmental, or other organizational entity
PERSON: named person or family
LOCATION: name of politically or geographically defined location (cities, provinces, countries, international regions, bodies of water, mountains, etc.)

2.3 Temporal Expressions (TIMEX tag element)

This subtask is for "absolute" temporal expressions only; explanation is provided in appendix B. The tagged tokens are categorized via the TYPE attribute as follows:

DATE: complete or partial date expression
TIME: complete or partial expression of time of day

2.4 Number Expressions (NUMEX tag element)

This subtask is for two useful types of numeric expressions, monetary expressions and percentages. The numbers may be expressed in either numeric or alphabetic form.

The task covers the complete expression, which is categorized via the TYPE attribute as follows:

MONEY: monetary expression
PERCENT: percentage

3 NOTATION RESERVED FOR USE IN THE ANSWER KEYS

3.1 Expressing Alternative Attribute Values

A vertical bar is being used to separate alternative TYPE attribute values in the answer key.

Alternative values will be given when the annotator does not have enough information to make a unique categorization, even considering the context and the annotator's knowledge of the world.

3.2 Expressing Optional Markup (STATUS Attribute)

When it is not certain that a string should be marked up, the annotator will include the STATUS attribute in the markup to indicate that the markup is optional. The only value of the STATUS attribute is "OPT." Examples of its possible use can be found in the appendices, such as in appendix B (holiday names).

3.3 Expressing Alternative String Boundaries (ALT Attribute)

The ALT attribute will be used when the tagged string contains one or more substrings that should be considered correct for the purposes of scoring the system response. Certain premodifiers ("a," "an," and "the") are automatically ignored by the scoring program (via the "configuration" file) and thus do not need to be specially marked in the key.

1 INTRODUCTION

1.1 Scope

The Named Entity task consists of three subtasks (entity names, temporal expressions, number expressions). The expressions to be annotated are "unique identifiers" of entities (organizations, persons, locations), times (dates, times), and quantities (monetary values, percentages).

For many text processing systems, such identifiers are recognized primarily using local pattern-matching techniques. The TEI (Text Encoding Initiative) Guidelines for Electronic Text Encoding and Interchange cover such identifiers (plus abbreviations) together in section 6.4 and explain that the identifiers comprise "textual features which it is often convenient to distinguish from their surrounding text. Names, dates and numbers are likely to be of particular importance to the scholar treating a text as source for a database; distinguishing such items from the surrounding text is however equally important to the scholar primarily interested in lexis."

The task is to identify all instances of the three types of expressions in each text in the test set and to subcategorize the expressions. The original texts contain some SGML tags already; the Named Entity task is to be performed within the text delimited by the TXT, HL, DATELINE, and DD tags. (Note, however, that the DD tag sometimes doesn't appear at all, sometimes appears once, and sometimes appears twice. When it appears twice, only the second instance is to be marked up for the Named Entity task.)

Lines within the <TXT> portion of the article that start with the "@" sign signify a table or other special line formatting within the text and should NOT be annotated. (However, such lines may also appear within the <HL> portion of the article, and these should be annotated.

The system must produce a single, unambiguous output for any relevant string in the text; thus, this evaluation is not based on a view of a pipelined system architecture in which Named Entity recognition would be completely handled as a preprocess to sentence and discourse analysis. The task requires that the system recognize what a string represents, not just its superficial appearance. Sometimes, the right answer is superficially apparent, as in the case of most, if not all, NUMEX expressions, and can be obtained by local pattern-matching techniques. In other cases, the right answer is not superficially apparent, as when a single capitalized word could represent the name of a location, person, or organization, and the answer may have to be obtained using techniques that draw information from a larger context or from reference lists.

The three subtasks correspond to three SGML tag elements: ENAMEX, TIMEX, and NUMEX. The subcategorization is captured by a SGML tag attribute called TYPE, which is defined to have a different set of possible values for each tag element. The markup is described in section 2, below.

1.2 Performance Evaluation

Scoring of this task will be done using the same kinds of metrics that are used for scoring template-filling (information extraction) tasks. For specific information on the scoring, refer to "MUC-6 Scoring System User's Manual," prepared for MUC-6 by SAIC.

Cumulative scores will be generated at several levels of description of the task, e.g.,

- across subtasks,
- for each subtask,
- for the subcategorization aspect of each subtask,
- for each part of the article that is included in the task (<HL>,<DATELINE>, <DD>, <TXT>).

2 TASK OVERVIEW

2.1 Markup Description

The output of the systems to be evaluated will be in the form of SGML text markup. The only insertions allowed during tagging are tags enclosed in angled brackets. No extra whitespace or carriage returns are to be inserted; otherwise, the offset count would change, which would adversely affect scoring.

The markup will have the following form:

Named Entity Task Definition

(Version 2.1, 27 JUL 95)

Updates since last version are indicated by a change bar ("|") in left margin.

1 INTRODUCTION . . . 2
1.1 Scope . . . 2
1.2 Performance Evaluation . . . 2
2 TASK OVERVIEW . . . 2
2.1 Markup Description . . . 2
2.2 Named Entities (ENAMEX tag element) . . . 3
2.3 Temporal Expressions (TIMEX tag element) . . . 3
2.4 Number Expressions (NUMEX tag element) . . . 3
3 NOTATION RESERVED FOR USE IN THE ANSWER KEYS . . . 3
3.1 Expressing Alternative Attribute Values . . . 3
3.2 Expressing Optional Markup (STATUS Attribute) . . . 3
3.3 Expressing Alternative String Boundaries (ALT Attribute) . . . 3
4 GUIDELINES FOR MARKUP OF EXCEPTIONAL CONSTRUCTIONS . . . 4
4.1 Expressions Involving Elision . . . 4
4.2 Effects of Tokenization Conventions . . . 4
4.3 Nested Expressions . . . 4
APPENDIX A. ENAMEX: SPECIFIC GUIDELINES . . . 5
APPENDIX B. TIMEX: SPECIFIC GUIDELINES . . . 10
APPENDIX C. NUMEX: SPECIFIC GUIDELINES . . . 12