

DESCRIPTION OF THE UNL/USL SYSTEM USED FOR MUC-3

Jitender S. Deogun
Department of Computer Science & Engineering
University of Nebraska - Lincoln
Lincoln, NE 68588-0115
deogun@fergvax.unl.edu
(402)-472-5033

Vijay V. Raghavan
Center for Advanced Computer Studies
University of Southwestern Louisiana
Lafayette, LA 70504-4330
raghavan@cacs.usl.edu
(318)-231-6603

BACKGROUND

The MUC-3 task consists of generating a database of filled templates with respect to messages that belong to a general topical domain. In particular, for the current phase, the message collection belongs to the domain of *terrorist activities*. On the one hand, a decision as to the relevance of a message to a specified class of terrorist events should be made. If relevant, a predefined set of facts are to be extracted and placed as fills for appropriate slots of the template(s) created for this message. If not relevant, a template having a '*' as the fill in all but one slot, is created (see Appendix A for details). Some aspects of the MUC-3 task are amenable to be solved by techniques typically employed in information retrieval (IR). These techniques are especially designed to be applicable to *any* domain. In contrast, there are other aspects of the problem that may require a great deal of language understanding, thus needing natural language processing (NLP) techniques. For the most part, NLP techniques may be considered domain dependent.

The primary thrust of our effort has been to design and implement a system that employs techniques typically found in IR literature, augmented by basic search techniques available in file management systems. An important goal (for the time being) is to ensure that the system is domain independent to the greatest extent possible. Consequently, certain slots which are not suitable to be handled by the chosen techniques are not filled.

In the context of the MUC-3 task, slots fall into one of four categories depending on the type of fill that are applicable to them. Our system is designed to handle slots whose values are from a *set-list*. More specifically we process TYPE OF INCIDENT, CATEGORY OF INCIDENT, PERPETRATOR: CONFIDENCE, PHYSICAL TARGET TYPE, HUMAN TARGET TYPE, INSTRUMENT: TYPE(S), EFFECT ON PHYSICAL TARGET(S) and EFFECT ON HUMAN TARGET(S). In addition, two slots whose fills are of *string* type are also processed. These are PERPETRATOR: ID OF ORG(S) and LOCATION OF INCIDENT.

As will be explained later, the system consists of an Indexing Module, a Learning Module, a Filtering Module, and a Template Filler Module. We had developed and experimentally validated indexing and learning techniques for use in the context of information retrieval and classification. These techniques were adapted to develop the indexing and learning modules for MUC-3 in addition

to the development and implementation of other modules. This site did not participate in either MUC-1 or MUC-2.

OVERVIEW OF THE SYSTEM

A popular strategy in IR is to formulate the problem of identifying items relevant to a subject area as one of conceptual categorization. The subject area(s) of interest is imagined as a concept or a class. Example items relevant to a certain concept are assumed to be given. Based on this information and using techniques for learning from examples, a concept characterization rule, that is optimal in a certain precise sense, is derived. In other words, retrieval of relevant items is actually viewed as a "recognition" problem.

Our system employs the above idea by mapping possible fill values of set-list type slots to *concepts of interest*. For example, in the context of the TYPE OF INCIDENT slot, fill values such as ARSON, MURDER, BOMBING, etc. are the concepts to be learned. Note that the question of whether concept ARSON is applicable to a message is equivalent to deciding whether a message belongs to the message class identified by the label ARSON. Thus, the template filling task and the decision of whether a message is relevant to MUC-3 task are investigated as problems requiring conceptual categorization.

For each concept that is considered by the system to be applicable to a message, the system also keeps track of the extent to which each of the paragraphs in the message contributed to this decision. Judicious use of this information enables various important activities such as the resolution of the "best" fill for a slot from among alternatives, the linking of the fills to templates when more than one template must be generated for the same message, and the filling of the two string type slots.

The general architecture of the system is presented in Figure 1. There are four major subsystems: Indexing Module, Learning Module, Filtering Module, and Template Filler Module. Each of these systems are outlined next.

Indexing Module

The function of the Indexing Module is to generate a representation for each message. A message is represented by a vector of weights. Each weight value either indicates the presence or absence of a *term* in the message or the importance of a term to the message. A term is either a single-order term or a high-order term (i.e., single terms or word combinations representing phrases).

For the assignment of single terms to messages, the indexing module from the SMART Retrieval System [1] is used. This module utilizes a stop list to filter out the common words and the "no-stemming" option is chosen. All terms that are assigned a weight larger than a threshold by this module are retained in the message representation vector.

For the purpose of phrase extraction, a modified version of the INDEX software, developed and implemented by Jones, et al. [2,3], is used. INDEX is used mainly to extract all possible substrings that are within certain minimum and maximum length specifications and are not substrings of other previously selected substrings. Several strategies for filtering these to identify "good" phrases are provided as a part of the software developed for the MUC-3 project.

Thus, each element of the vector representing a message corresponds to either a single term or a phrase. The phrase identification is expected to be important as a precision improving device. This module also generates the system vocabulary, which consists of all the distinct single terms and phrases used in representing the messages.

Learning Module

The function of the learning module is to derive the concept categorization rules for the various concepts of interest. Each rule is a vector of numeric weights, where the elements correspond to the terms in the system vocabulary.

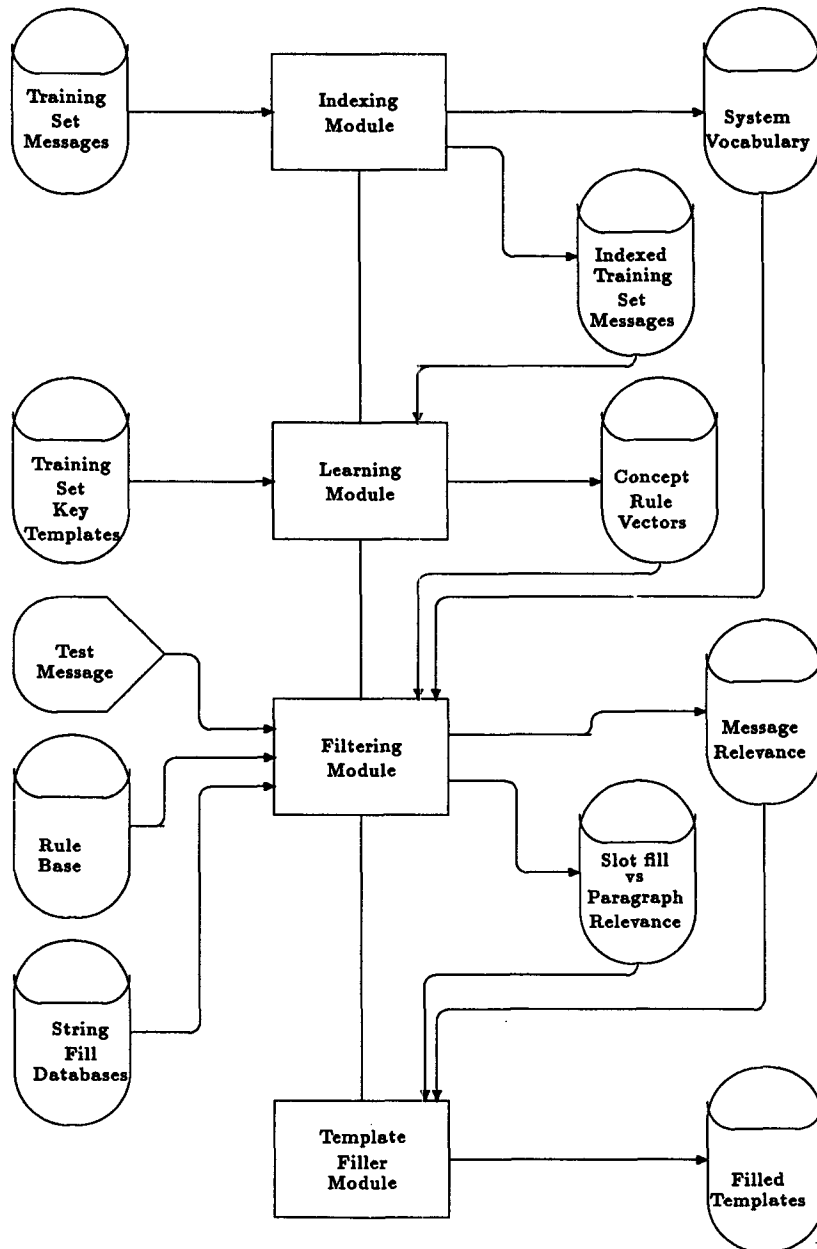


Figure 1: Block diagram of the system

This module also involves components for selecting a training set from the development set, identifying the concepts for which the training set has at least a minimum number of positive examples (i.e., the learnable concepts) and preparing the grid file, which shows for each message in the training set which of the learnable concepts are applicable. The source for this information is the set of key-templates manually generated for the 1300 messages in the development set.

The concept rule vectors are derived by employing the perceptron-learning algorithm [4]. The algorithm is simple and efficient. The procedure is incremental in that the rule can be updated as new examples become available. As long as a decision boundary exists, this algorithm is guaranteed to find one and terminate.

Usually, the decision boundary constructed is a hyper-plane. However, since the system vocabulary includes phrases, and phrases incorporate dependency information between single terms, our result is equivalent to constructing a non-linear boundary. In the terminology of the connectionist networks, we employ a single-layer, high-order perceptron. The single-layer option facilitates fast learning time, while the higher-order option enables the use of more powerful separation boundaries. Furthermore, the concept rule vectors are connectionist, rather than symbolic in nature. Such rules are more attractive when a large number of features are involved and when robustness against noisiness in features is crucial.

In addition to concepts associated with slot fills, another concept known as "optimal-query" is also derived. This rule vector distinguishes messages that are not relevant to MUC-3 task from those that generate at least one template. The system is set up in such a way that the training set of messages for deriving this optimal-query vector can be different from that used for the other concepts.

Filtering module

The Filtering Module is responsible for identifying concepts applicable to a set of test messages and deciding whether a message is relevant to the MUC-3 task. The major subsystems of this module are concerned with test message indexing, assessment of concept relevance and the evaluation of a rule base by means of an inference engine.

The test message indexing involves the determination of which of the single terms and phrases in the system vocabulary are contained in the message. This process generates a message vector that is matched against each of the concept rule vectors to determine the corresponding activation values. The distribution of the activation values for the test set of messages relative to each concept is analyzed to determine a threshold. A concept is considered relevant to a message if the corresponding activation value exceeds the threshold chosen for that concept. Depending on the concepts applicable to a message, the inference engine activates appropriate rules of the rule base, whose terminal symbols correspond to the various concepts acquired. The rule base expresses the requirements in terms of concept combinations that, when present in a message, imply that the message is relevant to MUC-3 task. The module also identifies for each message, the extent to which its paragraphs contributed to the activation values relative to the different concepts. This result is referred to as the concept vs. paragraph relevance vector.

For slots of string fill type, a database of possible fill values, grouped by slot name, is provided as input to this module. For each string in the database for which at least one match is found in the message, the paragraphs in which a match is found and the frequency of its occurrence in each paragraph is determined.

Template Filler Module

This module is responsible for generating one or more templates for each message determined to be relevant by the Filtering Module and filling the slots on the basis of concepts and string filled that are activated.

For each relevant message either the optimal-query concept is activated or one or more incident types are recognized along with a desired combination of concepts (or both). In the case exactly one incident type is recognized, for each of the other slots the following is performed. If several concepts

are activated for this slot and only one value is permitted, the one with highest activation value is chosen; otherwise, all values are filled.

In the case more than one incident type is activated, the system must decide, for each activated concept, to which incident type it is the closest. For this purpose, the concept versus paragraph relevance vector is used. This vector contains the contribution of the various paragraphs in a message to the activation value of the slot fill relative to this message. The paragraph relevance vector of an activated slot fill, say CIVILIAN (from HUMAN TARGET TYPE), is compared to the vector associated with each of the activated incident types, say KIDNAPPING and MURDER. The strength of this match is then used to decide whether the fill CIVILIAN will be used in the KIDNAPPING or the MURDER template.

If a message becomes relevant only due to optimal-query, then it enables other slots having activated fills to be filled even though no TYPE OF INCIDENT may have been activated.

SYSTEM WALKTHROUGH

The system walkthrough explains how the message TST1-MUC3-0099 is processed. The result obtained corresponds to the parameter settings used in our Option 4 (see report on UNL/USL: MUC-3 Test Results and Analysis). In this option, Training Set 2 is used for determining the rule vector for optimal-query and Training Set 3 is used for the other concepts. The threshold used for deciding whether a concept is activated is based on an analysis of the distribution of the activation values of this concept relative to the test set messages (threshold setting T1).

Table 1 shows a list of all set-list type fills and those that are actually learnable on the basis of training set 3. The concept rule vectors for each of these fills are constructed by using the indexing and the learning module. The test message is indexed and the dot product of its representation vector with each of the concept rule vectors is computed. The activation values so obtained are compared to the corresponding threshold values. Table 2 shows that, for the current message, the following five concepts are activated: BOMBING, TERRORIST ACT, TRANSPORT VEHICLE, SOME DAMAGE, and the optimal-query. These concepts activate the appropriate leaf nodes of the AND/OR tree associated with the rulebase shown in Table 4. This results in the root node getting the value "true" and therefore, this message is termed relevant. For the current testing, the rule base is defined with all the concept weights being either 0 or 1. The inference engine is, however, capable of handling any numeric weights between 0 and 1. The vector representation for each of the paragraphs in the message are also multiplied by the concept rule vectors to obtain the paragraph vs concept relevance vector (Table 3). This paragraph information is not useful in this case since neither several fills are activated for a slot for which only one fill is permitted nor is there an indication, in terms of INCIDENT TYPE activations, that multiple templates should be created.

For the two string fill slots the matching strings along with their occurrence frequency in the various paragraphs is shown in Table 5. The paragraph vector for BOMBING is found to match the paragraph vector of "POLICE" better (wrong decision!). All 3 incident locations have a positive activation value with BOMBING. Since the location slot permits multiple fills, all three may be retained. However, since "PRC" is not one of the South American countries, it is discarded.

The filled template for this message is shown in Table 6. This template most closely matches key-template that is numbered 2 (see Appendix H). The paragraph relevance vector matching technique needs to be refined as evidenced by the choice of "POLICE" as the perpetrator organization. Furthermore, template filler module should be refined to automatically determine and incorporate in the filling process various dependencies between template fills. For example, "POLICE" is inconsistent with CATEGORY OF INCIDENT being TERRORIST ACT.

By proper modification of the stop list used during phrase extraction, phrases such as NO INJURY could be extracted. The optimal-query vector identifies relevant passages fairly accurately. Careful detailed analysis of individual instances should lead to many ideas for improvement.

Concept	CID		Concept Name	Output file
concept1	1	✓	ARSON	slot.3
concept2	2		ARSON THREAT	slot.3
concept3	3	✓	MURDER	slot.3
concept4	4	✓	DEATH THREAT	slot.3
concept5	5	✓	BOMBING	slot.3
concept6	6		BOMB THREAT	slot.3
concept7	7	✓	KIDNAPPING	slot.3
concept8	8		KIDNAPPING THREAT	slot.3
concept9	9		HJACKING	slot.3
concept10	10		HJACKING THREAT	slot.3
concept11	11		ROBBERY	slot.3
concept12	12		ROBBERY THREAT	slot.3
concept13	13	✓	ATTACK	slot.3
concept14	14		ATTEMPTED ARSON	slot.3
concept15	15		ATTEMPTED MURDER	slot.3
concept16	16	✓	ATTEMPTED BOMBING	slot.3
concept17	17		ATTEMPTED KIDNAPPING	slot.3
concept18	18		ATTEMPTED HJACKING	slot.3
concept19	19		ATTEMPTED ROBBERY	slot.3
concept20	20	✓	TERRORIST ACT	slot.4
concept21	21		SABOTAGE	slot.4
concept22	22	✓	STATE-SPONSORED VIOLENCE	slot.4
concept23	23	✓	COMMERCIAL	slot.10
concept24	24	✓	COMMUNICATIONS	slot.10
concept25	25	✓	DIPLOMAT OFFICE OR RESIDENCE	slot.10
concept26	26		ENERGY	slot.10
concept27	27		FINANCIAL	slot.10
concept28	28		GOVERNMENT OFFICE OR RESIDENCE	slot.10
concept29	29		NONGOVERNMENT	slot.10
concept30	30		ORGANIZATION	slot.10
concept31	31	✓	TRANSPORT VEHICLE	slot.10
concept32	32		TRANSPORTATION FACILITY	slot.10
concept33	33	✓	OTHER	slot.10
concept34	34	✓	CIVILIAN	slot.13
concept35	35		DIPLOMAT	slot.13
concept36	36	✓	GOVERNMENT OFFICIAL	slot.13
concept37	37	✓	FORMER GOVERNMENT OFFICIAL	slot.13
concept38	38		FORMER ACTIVE MILITARY	slot.13
concept39	39	✓	LEGAL OR JUDICIAL	slot.13
concept40	40		NONGOVERNMENT POLITICIAN	slot.13
concept41	41	✓	LAW ENFORCEMENT	slot.13
concept62	62	✓	REPORT AS FACT	slot.7
concept63	63	✓	CLAIMED OR ADMITTED	slot.7
concept64	64		CLAIMED OR ADMITTED BY GOVERNMENT	slot.7
concept65	65	✓	SUSPECTED OR ACCUSED	slot.7
concept66	66	✓	SUSPECTED OR ACCUSED BY AUTHORITIES	slot.7
concept67	67		POSSIBLE	slot.7
concept68	68	✓	GUN	slot.15
concept69	69	✓	MACHINE GUN	slot.15
concept70	70		RIFLE	slot.15
concept71	71		MORTAR	slot.15
concept72	72		EXPLOSIVE	slot.15
concept73	73		BOMB	slot.15
concept74	74		GRENADE	slot.15
concept75	74		FIRE	slot.15
concept76	76		TORTURE	slot.15
concept77	77	✓	DESTROYED	slot.17
concept78	78	✓	SOME DAMAGE	slot.17
concept79	79	✓	NO DAMAGE	slot.17
concept80	80	✓	INJURY	slot.18
concept81	81		DEATH	slot.18
concept82	82		NO DAMAGE	slot.18
concept83	83		NO INJURY	slot.18
concept84	84	✓	NO INJURY OR DEATH	slot.18
concept85	85		NO RESIGNATION	slot.18
concept86	86		RESIGNATION	slot.18
concept87	87	✓	OPTIMAL_QUERY	

Table 1: List of Set List Slot Fills

CID	CEV	Cutoff	Concept Name
1	-536.0	-140.0	
3	-175.0	+47.0	
4	-599.0	-130.0	
5	+159.0	-178.0	BOMBING
7	-535.0	-192.0	
13	-321.0	-95.0	
16	-555.0	-138.0	
20	+333.0	+51.0	TERRORIST ACT
22	-455.0	-93.0	
23	-432.0	-131.0	
24	-517.0	-149.0	
25	-347.0	-140.0	
31	+10.0	-116.0	TRANSPORT VEHICLE
33	-538.0	-159.0	
34	-48.0	-29.0	
36	-379.0	-148.0	
37	-477.0	-170.0	
39	-574.0	-142.0	
41	-460.0	-158.0	
62	-45.0	-22.0	
63	-410.0	-142.0	
65	-236.0	-148.0	
66	-241.0	-122.0	
68	-384.0	-160.0	
69	-475.0	-156.0	
77	-441.0	-155.0	
78	+144.0	-122.0	SOME DAMAGE
79	-516.0	-148.0	
80	-71.0	-59.0	
84	-232.0	-169.0	
87	+573.0	+0.00	OPTIMAL_QUERY

Table 2: Concept activation and cut off values for TST1-MUC3-0099

CID	CEV	Paragraph Information									Cutoff
1	-536.0	1: -107.0	4: -185.0	5: -80.0	6: -91.0	7: -169.0	8: -41.0	2: -164.0	3: -63.0	-140.0	
3	-175.0	1: -28.0	4: -7.0	5: -20.0	6: 32.0	7: -23.0	8: 17.0	2: -83.0	3: -59.0	+47.0	
4	-599.0	1: -100.0	4: -165.0	5: -88.0	6: -93.0	7: -162.0	8: -50.0	2: -177.0	3: -93.0	-130.0	
5	+159.0	1: 28.0	4: -31.0	5: 28.0	6: -31.0	7: 1.0	8: -13.0	2: 95.0	3: 52.0	-178.0	
7	-535.0	1: -94.0	4: -150.0	5: -79.0	6: -84.0	7: -147.0	8: -37.0	2: -154.0	3: -77.0	-192.0	
13	-321.0	1: -22.0	4: -101.0	5: -52.0	6: -75.0	7: -92.0	8: -32.0	2: -116.0	3: -27.0	-95.0	
16	-555.0	4: -171.0	8: -50.0	2: -118.0	7: -132.0	3: -85.0	1: -92.0	5: -71.0	6: -84.0	-138.0	
20	+333.0	1: 62.0	4: 113.0	5: 52.0	6: 30.0	7: 76.0	8: 16.0	2: 117.0	3: 43.0	+51.0	
22	-455.0	1: -77.0	4: -125.0	5: -56.0	6: -30.0	7: -102.0	2: -167.0	3: -71.0	8: -16.0	-93.0	
23	-432.0	4: -141.0	8: -30.0	2: -139.0	7: -149.0	1: -50.0	5: -53.0	3: -58.0	6: -74.0	-131.0	
24	-517.0	1: -81.0	4: -153.0	5: -88.0	6: -100.0	7: -171.0	8: -50.0	2: -168.0	3: -64.0	-149.0	
25	-347.0	1: -73.0	4: -136.0	5: -83.0	6: -80.0	7: -53.0	8: -31.0	2: 25.0	3: -49.0	-140.0	
31	+10.0	1: -5.0	4: -38.0	5: 77.0	6: -45.0	7: -35.0	8: -19.0	2: 13.0	3: 15.0	-116.0	
33	-538.0	1: -99.0	4: -192.0	5: -63.0	6: -81.0	7: -148.0	8: -34.0	2: -173.0	3: -80.0	-159.0	
34	-48.0	1: 13.0	4: 0.0	5: 18.0	6: 18.0	7: -28.0	2: 8.0	8: 15.0	3: -8.0	-29.0	
36	-379.0	1: -59.0	4: -81.0	5: -59.0	6: -43.0	7: -97.0	8: -32.0	2: -148.0	3: -81.0	-148.0	
37	-477.0	1: -87.0	4: -95.0	5: -70.0	6: -69.0	7: -158.0	8: -14.0	2: -138.0	3: -82.0	-170.0	
39	-574.0	1: -89.0	4: -163.0	5: -90.0	6: -95.0	7: -166.0	8: -50.0	2: -158.0	3: -88.0	-142.0	
41	-460.0	4: -110.0	8: -44.0	2: -152.0	7: -148.0	3: -80.0	1: -60.0	5: -77.0	6: -65.0	-159.0	
62	-45.0	1: -12.0	4: -2.0	5: -11.0	6: -13.0	7: 9.0	2: 12.0	3: 8.0	8: 15.0	-22.0	
63	-410.0	1: -76.0	4: -113.0	5: -42.0	6: -63.0	7: -110.0	8: -40.0	2: -151.0	3: -48.0	-142.0	
65	-236.0	1: -29.0	4: -56.0	5: -56.0	6: -37.0	7: -75.0	8: -10.0	2: -101.0	3: -43.0	-148.0	
66	-241.0	1: -30.0	4: -20.0	5: -10.0	6: -29.0	7: -55.0	8: -22.0	2: -71.0	3: -71.0	-122.0	
68	-384.0	1: -58.0	4: -91.0	5: -46.0	6: -49.0	7: -95.0	8: -14.0	2: -97.0	3: -63.0	-160.0	
69	-475.0	1: -81.0	4: -92.0	5: -37.0	6: -26.0	7: -116.0	8: -36.0	2: -164.0	3: -100.0	-156.0	
77	-441.0	1: -76.0	4: -148.0	5: -76.0	6: -90.0	7: -152.0	8: -50.0	2: -134.0	3: -21.0	-155.0	
78	+144.0	1: 21.0	4: -53.0	5: 31.0	6: -35.0	7: 4.0	8: -15.0	2: 102.0	3: 53.0	-122.0	
79	-516.0	1: -82.0	4: -176.0	5: -95.0	6: -100.0	7: -158.0	8: -50.0	2: -90.0	3: -58.0	-148.0	
80	-71.0	1: 8.0	4: -63.0	5: 27.0	6: -31.0	7: -69.0	2: 4.0	8: -29.0	3: -22.0	-59.0	
84	-232.0	1: -30.0	4: -120.0	5: -61.0	6: -79.0	7: -81.0	8: -36.0	2: -32.0	3: 26.0	-169.0	
87	+573.0	4: 196.0	2: 273.0	1: 144.0	5: 51.0	6: 3.0	7: 18.0	8: 117.0	3: 177.0	+0.0	

Table 3: Paragraph relevance vectors for concepts for TST1-MUC3-0099

MUC3 \Leftarrow **RULE** (1.0) \vee **concept87** (1.0)
RULE \Leftarrow **INCIDENT_TYPE** (1.0) \wedge **location** (1.0)
INCIDENT_TYPE \Leftarrow **ATTACK** (1.0) \vee **ARSON** (1.0) \vee **BOMBING** (1.0) \vee **DEATH.THREAT** (1.0) \vee **MURDER** (1.0)
 \vee **KIDNAPPING** (1.0) \vee **ATTEMPTED_BOMB** (1.0)
ATTACK \Leftarrow **concept13** (1.0) \wedge **COMBINE_CONCEPT** (1.0)
ARSON \Leftarrow **concept1** (1.0) \wedge **COMBINE_CONCEPT** (1.0)
BOMBING \Leftarrow **concept5** (1.0) \wedge **COMBINE_CONCEPT** (1.0)
DEATH.THREAT \Leftarrow **concept4** (1.0) \wedge **COMBINE_CONCEPT** (1.0)
MURDER \Leftarrow **concept3** (1.0) \wedge **COMBINE_CONCEPT** (1.0)
KIDNAPPING \Leftarrow **concept7** (1.0) \wedge **COMBINE_CONCEPT** (1.0)
ATTEMPTED_BOMB \Leftarrow **concept16** (1.0) \wedge **COMBINE_CONCEPT** (1.0)
COMBINE_CONCEPT \Leftarrow **EFFECT** (1.0) \vee **INSTRUMENT** (1.0) **TARGET** (1.0) \vee **organ** (1.0) \vee **CATEGORY**
(1.0) \vee **CONFIDENCE** (1.0)
EFFECT \Leftarrow **HUM_EFFECT** (1.0) \vee **PHY_EFFECT** (1.0)
HUM_EFFECT \Leftarrow **concept80** (1.0) \vee **concept84** (1.0)
PHY_EFFECT \Leftarrow **concept77** (1.0) \vee **concept78** (1.0) \vee **concept79** (1.0)
INSTRUMENT \Leftarrow **concept68** (1.0) \vee **concept69** (1.0)
TARGET \Leftarrow **PHY_TARGET** (1.0) \vee **HUM_TARGET** (1.0)
HUM_TARGET \Leftarrow **concept34** (1.0) \vee **concept36** (1.0) \vee **concept37** (1.0) \vee **concept39** (1.0) \vee **concept41** (1.0)
PHY_TARGET \Leftarrow **concept23** (1.0) \vee **concept24** (1.0) \vee **concept25** (1.0) \vee **concept31** (1.0) \vee **concept33** (1.0)
CATEGORY \Leftarrow **concept20** (1.0) \vee **concept22** (1.0)
CONFIDENCE \Leftarrow **concept62** (1.0) \vee **concept63** (1.0) \vee **concept65** (1.0) \vee **concept66** (1.0)

Note: The highlighted predicates have true values. Lower case predicates are terminals, whereas the ones in upper case are non-terminals.

Table 4: Relevance Judgement of Rule Set for TST1-MUC3-0099

The organizations in TST1-MUC3-0099					
#TST1-MUC3-0099					
POLICE	1:1	3:1			
SHINING PATH	4:1	5:1	6:1	7:1	8:1
The locations in TST1-MUC3-0099					
#TST1-MUC3-0099					
LIMA	1:1	2:1	8:1		
PRC	2:1				
SAN ISIDRO	2:1				

Table 5: Paragraph occurrence vectors for string slot fills for TST1-MUC3-0099

0.	MESSAGE ID	TST1-MUC3-0099
1.	TEMPLATE ID	1
2.	DATE OF INCIDENT	-
3.	TYPE OF INCIDENT	BOMBING
4.	CATEGORY OF INCIDENT	TERRORIST ACT
5.	PERPETRATOR: ID OF INDIV(S)	-
6.	PERPETRATOR: ID OF ORG(S)	"POLICE"
7.	PERPETRATOR: CONFIDENCE	-
8.	PHYSICAL TARGET: ID(S)	-
9.	PHYSICAL TARGET: TOTAL NUM	-
10.	PHYSICAL TARGET: TYPE(S)	TRANSPORT VEHICLE
11.	HUMAN TARGET: ID(S)	-
12.	HUMAN TARGET: TOTAL NUM	-
13.	HUMAN TARGET: TYPE(S)	-
14.	TARGET: FOREIGN NATION(S)	-
15.	INSTRUMENT: TYPE(S)	*
16.	LOCATION OF INCIDENT	PERU: SAN ISIDRO PERU: LIMA
17.	EFFECT ON PHYSICAL TARGET(S)	SOME DAMAGE
18.	EFFECT ON HUMAN TARGET(S)	-

Table 6: The filled template for TST1-MUC3-0099

REFERENCES

1. Buckley, C. (1985), "Implementation of the SMART information retrieval system", TR 85 - 686, Dept. of Computer Science, Cornell University, Ithaca, NY.
2. L. P. Jones, E. W. Gassie. Jr., S. Radhakrishnan, "INDEX : Statistical basis for an automatic conceptual phrase-index system", Journal of American Society for Information Science, Vol. 41, pp. 87 - 91, 1990.
3. L. P. Jones, E. W. Gassie. Jr., S. Radhakrishnan, "PORTREP : A portable repeated string finder", Software Practice and Experience, Vol. 19, pp. 63 - 77, 1989.
4. Duda, R. O. and Hart, P. E. (1973), Pattern Classification and Scene Analysis, Wiley, NY.

PART IV: OTHER CONTRIBUTED PAPERS

The papers in this section provide two general perspectives on MUC-3 that came out of the evaluation task. The first paper describes an experimental system based on a statistical text categorization technique. The results of testing that system on the MUC-3 task help in assessing the difficulty of the task and the appropriateness of text categorization as an element of a complete information extraction system. The second one is a joint paper prepared by representatives of seven of the participating sites on the subject of discourse analysis as it pertains to MUC-3. The desire to offer this paper arose from the common perception of the discourse handling demands placed on the systems by the MUC-3 corpus and task.