# Generating Single and Multi-Document Summaries with GISTEXTER

**Sanda M. Harabagiu***, **Finley Lăcătuşu**

*Language Computer Corporation
Dallas TX 75206 USA
sanda@languagecomputer.com

## Abstract

This paper presents the techniques implemented in GISTEXTER for producing *extracts* and *abstracts* from both single and multiple documents. These techniques promote the belief that highly coherent summaries may be generated when using textual information identified by the Information Extraction technology. The results of GISTEXTER in the DUC-2002 evaluations account for the advantages of using the techniques presented in this paper.

## 1. Introduction

One way of tackling the current textual information overload is by relying on summaries of either single documents or of sets of documents that share the same category or cover the same topic from multiple perspectives. Summaries compress the information content available in a long text or a text collection by producing a much shorter text that can be read and interpreted rapidly. At the core of automatic summarization techniques that produce coherent summaries stays the methodology of identifying in the original documents the relevant information that should be included in the summary. Similarly, Information Extraction (IE) is a technology that targets the identification of topic-related information in free text and translates it into database entries. Typically, IE systems extract around 10% if a document textual content (cf. (Hobbs and et al.1997)). This represents a compression ratio that qualifies extraction techniques for multi-document summarization. Our automatic summarization system, called GISTEXTER builds on this observation.

To further progress in summarization and enable researchers to participate in large-scale experiments, the National Institute of Standards and Technology (NIST) has initiated in 2001 an evaluation in the area of text summarization called the Document Understanding Conference (DUC)[1]. For DUC-2002 NIST produced 59 document sets as test data. For this purpose NIST used the TREC disks employed in the question-answering track in TREC-9. Specifically these include articles from *Wall Street Journal (1987-1992)*, *AP newswire (1989-1990)*, *San Jose Mercury News (1991)*, *Financial Times (1991-1994)*, *LA Times* and *FBIS records*. Each set had between 5 and 15 documents, with an average of 10 documents. The documents were at least 10 sentences long, but there was no maximum length. Additionally, NIST classified the 59 documents sets in the categories listed in Figure 1. For each document in the test data, the sentences were tagged by NIST.

*Category 1: documents about a single natural disaster and created within at most a seven day window.*

*Category 2: documents about a single event in any domain created within at most a seven day window.*

*Category 3: documents about multiple distinct events of a single type (no limit on the time window)*

*Category 4: documents that present biographical information mainly about a single individual*

Figure 1: Definitions of document set categories.

Three different tasks were evaluated in DUC-2002:

1. Fully automatic summarization of a single newswire/newspaper document. Given a single document, a generic abstract of the document with a length of approximately 100 words or less was required. The abstracts were composed entirely of complete sentences.

2. Fully automatic summarization of multiple newswire/newspaper documents on a single subject by generating *document extracts*. Given a set of documents, 2 generic sentence extracts of the entire set with lengths of approximately 400 and 200 (whitespace-delimited tokens) or less were required. Each such extract consisted of some subset of the ”sentences” predefined by NIST in the sentence-separated document set. Each predefined sentence had be used in its entirety or not at all in constructing an extract.

3. Fully automatic summarization of multiple newswire/newspaper documents on a single subject by generating *document abstracts*. Given a set of documents, we had to create 4 generic abstracts of the entire set with lengths of approximately 200, 100, 50, and 10 words (whitespace-delimited tokens) or less. The 200, 100, and 50-word abstracts had to be composed entirely of complete sentences. The 10-word abstract took the form of a headline.

To train summarization systems, NIST provided 30 document sets with assorted, human-generated abstracts for single and multiple documents, prepared for the DUC-2001, as well as combined test and training data from

[1] DUC is part of a Defense Advanced Research Projects Agency (DARPA) program, Translingual Information Detection, Extraction, and Summarization (TIDES), which specifically calls for major advances in summarization technology, both in English and from other languages to English (cross-language summarization)

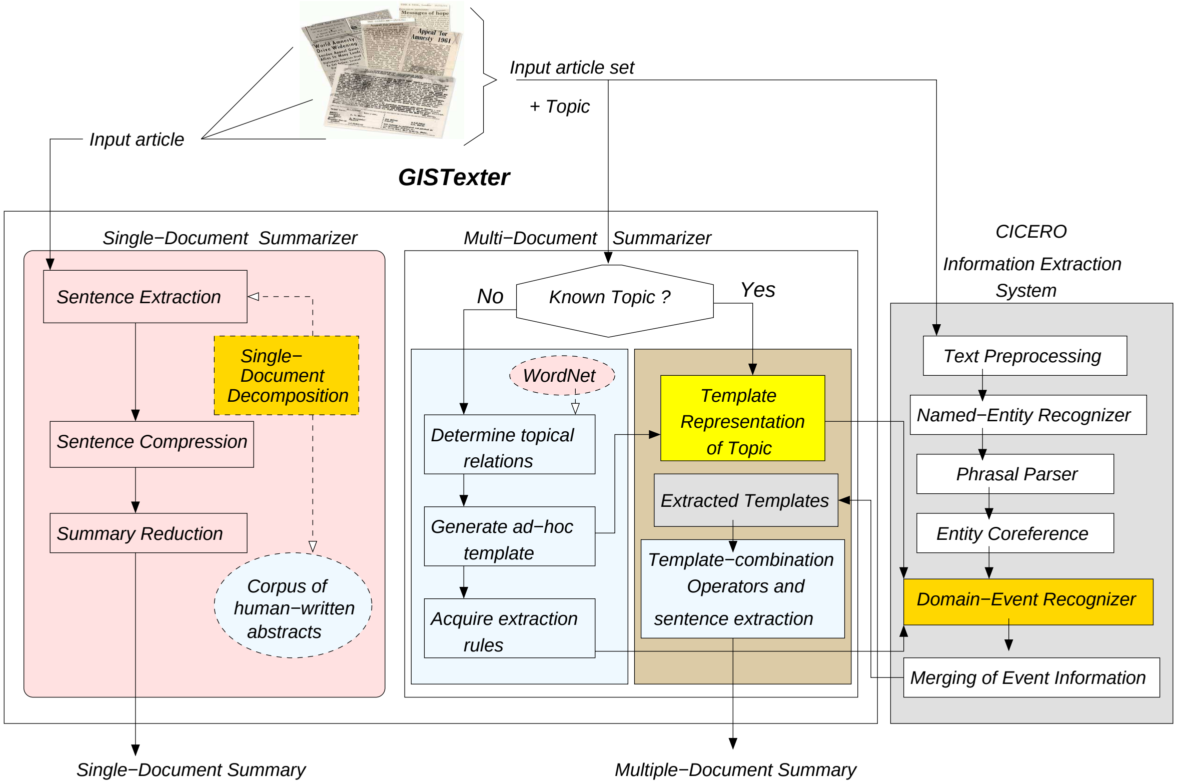


Figure 2: Architecture of GISTEXTER

DUC-2001. For single document summaries there were 2 categories of evaluation: that done by humans (mostly at NIST), and that done automatically (outside of NIST). For multi-document summarization, the plan was only to have human evaluation. Human evaluation was done at NIST using the same personnel who created the reference data. These people did pairwise comparisons of the reference summaries to the system-generated summaries, other reference summaries, and baseline summaries.

The rest of the paper is organized as follows. Section 2 presents the architecture of GISTEXTER, our single-document and multi-document summarization system. Section 3 presents the IE-based multi-document summarization producing extracts whereas Section 4 presents ad-hoc extraction techniques for multi-document summarization. Section 5 reports and discusses the experimental results we obtained in DUC-2002 and Section 6 summarizes the conclusions.

## 2. The architecture of GISTEXTER

GISTEXTER is a summarization system implemented for the evaluations of the Document Understanding Conferences (DUCs)[2]. The architecture of the system is shown in Figure 2. Input to the system is either a single document or a collection of documents sharing the same topic. When a summary of a single document is sought, GISTEXTER first extracts the key sentences, similarly to most single-document summarizers. The *sentence extraction* function is learned, using the technique of *single-document decomposition*. This technique analyzes the features of human-written abstracts of single documents. In the second stage, to further filter out un-necessary information, the extracted sentences are compressed. In the final stage a *summary reduction* is performed, to trim the whole summary to the length of 100 words. Figure 3 illustrates a single-document summary produced by GISTEXTER in DUC-2002.

*Hurricane Gilbert, packing 110 mph winds and torrential rain, moved over this capital city today, 09/12/1988, after skirting Puerto Rico, Haiti and the Dominican Republic.*
*There were no immediate reports of casualties.*
*Forecasters say Gilbert was expected to lash Jamaica throughout the day and was on track to later strike the Cayman Islands, a small British dependency northwest of Jamaica.*
*The Associated Press' Caribbean headquarters in San Juan, was unable to get phone calls through to Kingston, where high winds and heavy rain preceding the storm drenched the capital overnight, toppling trees, causing local flooding and littering streets with branches.*

Figure 3: Single-document summary produced by GISTEXTER.

When multi-document summaries need to be created, the processing takes additionally into account the topic of the document set. Sometimes the topic is well-known and may be already implemented in Information Extraction (IE) systems. In this case an IE system identifies all the information that needs to be used in the multi-document summary. Other times the topic is completely new, and the summary is generated by modeling the topic in an ad-hoc manner.

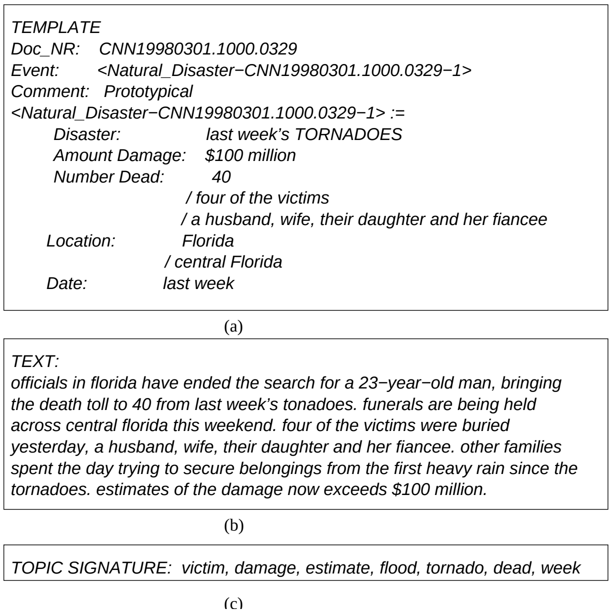
GISTEXTER produces multi-document summaries by relying on the output of the CICERO IE system[3]. CICERO,

[2] See http://www-nlpir.nist.gov/projects/duc/

[3] CICERO is an ARDA-sponsored on-going project that studies the effects of incorporating world knowledge into IE systems. CICERO is being developed at Language Computer Corporation.

as reported in (Surdeanu and Harabagiu 2002) produces unsurpassed quality of extraction because it combines the role of linguistic extraction patterns with coreference knowledge. For multi-document summarization, this means that the templates generated by CICERO are easily mapped into text snippets from the texts, in which pronouns and other anaphoric expressions are resolved. These text snippets can be used to generate coherent, informative multi-document summaries.

To extract information from a set of documents, CICERO needs to have a template representation of the topic. Topics can be represented as a set of inter-related concepts, implemented as a frame having slots and fillers. In the Information Extraction technology, such frames are called *templates* and are populated with information related to the salient facts reported in documents and extracted by the IE systems. For example, if the topic is *"natural disasters"*, Figure 4 illustrates a template populated with information extracted from the text illustrated in Figure 4(b). An alternative representation of a topic was proposed in (Lin and Hovy 2000), with the goal of modeling the minimum amount of knowledge required to effectively identify concepts related to a topic. This representation, called *topic signature*, associates a target concept (i.e. the topic) with a vector of related terms (i.e. the signature). Each $term_i$ from the signature has an associated weight $w_i$. (Lin and Hovy 2000) report on an automatic method of signature term extraction and weight estimation. Figure 4(c) illustrates the signature terms for the natural disasters topic, obtained with the method reported in (Lin and Hovy 2000).



```
TEMPLATE
Doc_NR:    CNN19980301.1000.0329
Event:       <Natural_Disaster−CNN19980301.1000.0329−1>
Comment:   Prototypical
<Natural_Disaster−CNN19980301.1000.0329−1> :=
        Disaster:              last week's TORNADOES
        Amount Damage:    $100 million
        Number Dead:          40
                                    / four of the victims
                                   / a husband, wife, their daughter and her fiancee
      Location:            Florida
                               / central Florida
      Date:                last week
```

(a)

*TEXT:*
*officials in florida have ended the search for a 23−year−old man, bringing the death toll to 40 from last week's tonadoes. funerals are being held across central florida this weekend. four of the victims were buried yesterday, a husband, wife, their daughter and her fiancee. other families spent the day trying to secure belongings from the first heavy rain since the tornadoes. estimates of the damage now exceeds $100 million.*

(b)

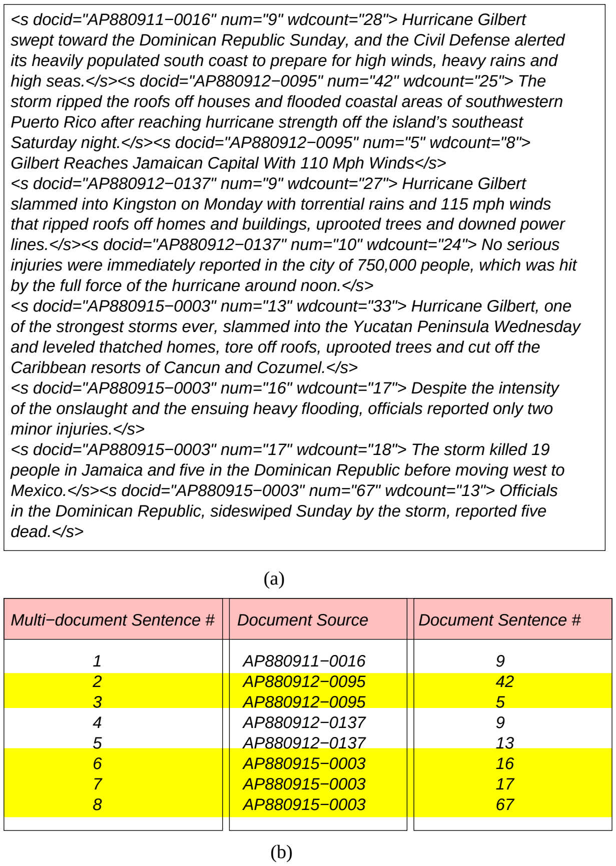
*TOPIC SIGNATURE: victim, damage, estimate, flood, tornado, dead, week*

(c)

Figure 4: (a) Template representation of the "natural disasters" topic; (b) Text containing information about the topic; (c) Topic signature for "natural disasters".

The template slots are filled whenever textual information relevant for the topic is identified. To recognize each topic-relevant event and entity, CICERO first pre-processes the text, by tokenizing the article and recognizing the part-of-speech and attributes of each word against a rich dictionary structure. Next, all names from the article are categorized by a named entity recognizer which tags *Red Cross* as an *Organization* and *Florida* as a *Location*. A phrasal parser brackets all noun and verb phrases, to enable the recognition of linguistic patterns that relate to the topic. Since anaphoric expressions are often used, before matching the text against linguistic patterns, coreference resolution takes place.

Linguistic patterns are matched to identify the topic-relevant information. For example, for the topic of "natural disasters", the rule $[$*Casualty-expression* $\{$*to*$|$*from*$\}$ *$Number* $\{$*from*$|$*because-of*$\}$ *Disaster-word*$]$ is matched against the snippet "the death toll to 40 from last week's tornado" in the text from Figure 4(b). Other extraction patterns are matched against the text and populate the rest of the template illustrated in Figure 4(a). CICERO extracts all the templates from the article collection and keeps mappings from the template slots the the text snippets containing information that fills the slots. These text snippets are indicators of the summary content. Additionally, reference resolution contributes to resolving the order of the sentences extracted from different documents. Since extracts are generated by selecting sentences marked-up by NIST in the documents, the summaries contain the SGML mark-up as well. For example Figure 5(a) illustrates the 200-word long multi-document summaries generated by GISTEXTER for a collection of articles dealing with "natural disasters".



*<s docid="AP880911−0016" num="9" wdcount="28"> Hurricane Gilbert swept toward the Dominican Republic Sunday, and the Civil Defense alerted its heavily populated south coast to prepare for high winds, heavy rains and high seas.</s><s docid="AP880912−0095" num="42" wdcount="25"> The storm ripped the roofs off houses and flooded coastal areas of southwestern Puerto Rico after reaching hurricane strength off the island's southeast Saturday night.</s><s docid="AP880912−0095" num="5" wdcount="8"> Gilbert Reaches Jamaican Capital With 110 Mph Winds</s>*
*<s docid="AP880912−0137" num="9" wdcount="27"> Hurricane Gilbert slammed into Kingston on Monday with torrential rains and 115 mph winds that ripped roofs off homes and buildings, uprooted trees and downed power lines.</s><s docid="AP880912−0137" num="10" wdcount="24"> No serious injuries were immediately reported in the city of 750,000 people, which was hit by the full force of the hurricane around noon.</s>*
*<s docid="AP880915−0003" num="13" wdcount="33"> Hurricane Gilbert, one of the strongest storms ever, slammed into the Yucatan Peninsula Wednesday and leveled thatched homes, tore off roofs, uprooted trees and cut off the Caribbean resorts of Cancun and Cozumel.</s>*
*<s docid="AP880915−0003" num="16" wdcount="17"> Despite the intensity of the onslaught and the ensuing heavy flooding, officials reported only two minor injuries.</s>*
*<s docid="AP880915−0003" num="17" wdcount="18"> The storm killed 19 people in Jamaica and five in the Dominican Republic before moving west to Mexico.</s><s docid="AP880915−0003" num="67" wdcount="13"> Officials in the Dominican Republic, sideswiped Sunday by the storm, reported five dead.</s>*

(a)

| *Multi−document Sentence #* | *Document Source* | *Document Sentence #* |
|---|---|---|
| 1 | AP880911−0016 | 9 |
| 2 | AP880912−0095 | 42 |
| 3 | AP880912−0095 | 5 |
| 4 | AP880912−0137 | 9 |
| 5 | AP880912−0137 | 13 |
| 6 | AP880915−0003 | 16 |
| 7 | AP880915−0003 | 17 |
| 8 | AP880915−0003 | 67 |

(b)

Figure 5: Multiple-document summary produced by GISTEXTER: (a) the 200-word extract and (b) the document-source table.

The SGML mark-up illustrated in Figure 5(a) contains three fields: *docid* indicating the document id of the source;

*num* indicating the sentence number on the source document and *wdcount* giving the length in words or tokens of the sentence. The mark-up in the running text of the summary contribute to mapping the order of the sentences in the summary to the sentence order in their original documents. For example, for the summary represented in Figure 5(a) a document-source table, as illustrated in Figure 5(b) is generated, showing both the source of each of the eight sentences as well as their respective order in the source documents. The sentences originated on four different documents, and except for the sentences extracted from document AP880912-0137, they follow the order from the source documents. The multi-document extracts generated by GISTEXTER are coherent because they rely on the relevant information identified by the IE system when the topic is known. However, out of the 59 topics covered in the document sets, 19 were not encoded in the CICERO IE system, thus they were considered new topics. Some of these topics are listed in Figure 6.

*McDonald's in Yugoslavia, Seoul, Soviet Union, China*
*Famous Allied Checkpoint Dividing East And West Berlin Removed*
*German Reunification*
*Dog Shows*
*The motion picture industry's most coveted award, Oscar*
*Iraq Invades Kuwait*
*Kashmir: A Tourist Paradise Becomes a War Zone*

Figure 6: Examples of new topics evaluated in DUC-2002.

Whenever the topic of the collection of documents has not been previously encoded in the CICERO IE system and no template representation of the topic exists, we need to perform some additional processing to gist the missing information. Thus we need to generate in an ad-hoc manner: (1) the template and (2) the extraction rules that enable CICERO to identify the relevant information. To this end, we have developed a methodology for generating an ad-hoc template based on the topical relations that can be identified from WordNet (Miller 1995). When the template is known, several possible methods of acquiring extraction rules can be applied, e.g. the methods reported in (**?**) (Riloff and Jones 1999) or (Harabagiu and Maioarano 2000). For GISTEXTER, we applied the techniques reported in (Harabagiu and Maioarano 2000).

With an ad-hoc template available, CICERO's domain-event recognizer acts in the same way as for topics that are encoded in the IE system. Moreover, entity coreference takes place for new topics also, since the coreference methods implemented in CICERO are topic-independent. The quality of the extraction is not be as good as in the case of previously studied topics because additional semantic knowledge is required to correctly merge incomplete templates. Nevertheless, for multi-document summarization, the extraction quality for ad-hoc templates is reasonable, as it determines acceptably coherent summaries. Example of multiple-document summaries produced by GISTEXTER for a new topic, namely the *"German Reunification"*, are illustrated in Figure 7.

Similar ad-hoc templates were also generated for document sets covering biographies of celebrities. Figure 8 lists some the focus of some of the biographies generated as multi-document extracts.

```
<s docid="WSJ890922−0113" num="9" wdcount="30"> The mass emigration
of thousands of disaffected East Germans has rekindled reunification talk in
West Germany, where some legislators plan to begin exploring the possiblity
of reuniting the two Germanys.</s>
<s docid="AP891111−0064" num="10" wdcount="32"> The lifting of travel
restrictions by East Germany on Thursday has breathed new life into the idea
of a single German state, drawing expressions of support from the Bush
administration and others.</s>
<s docid="AP891212−0062" num="12" wdcount="23"> The Communist Party
today admitted that East Germany's socialist system has failed, and expressed
support for a type of partnership with West Germany.</s>
<s docid="AP900130−0202" num="13" wdcount="22"> President Mikhail S.
Gorbachev met Tuesday with East German Premier Hans Modrow and
appeared to be more open toward eventual German reunification.</s>
<s docid="AP900210−0106" num="10" wdcount="23"> About 20,000 East
Germans, many carrying West German flags, demonstrated Saturday for speedy
German reunification, the official East German news agency ADN said.</s>
<s docid="LA021290−0043" num="9" wdcount="23"> West German Chancellor
Helmut Kohl declared on his return from Moscow on Sunday that "the way is
now free" for German reunification .</s>
<s docid="AP900215−0013" num="14" wdcount="48"> He was one of several
experts in German history discussing the implications of reunification following
the decision Tuesday by the four World War II allies _ the United States, Britain,
France and the Soviet Union _ to accept the reunification of a Germany they
divided 45 years ago.</s>
```

Figure 7: Multiple-document summary produced by GISTEXTER for the "mad-cow disease" topic: a 200-word text extract.

*Sakharov, the Nobel Peace Prize winner*
*Lucille Ball*
*Sam Walton*
*Erich Honecker, the former GDR head of state*
*Leonard Bernstein, pianist, composer, conductor, teacher*
*Margaret Thatcher, the first female prime minister in Europe*

Figure 8: Some biographical profiles evaluated in DUC-2002.

Finally, from extracts GISTEXTER generates abstracts of 10-, 59-, 100- and 200-word length by resolving the temporal expressions to absolute expressions and then compressing sentences to cover only the snippets identified by the IE system.

## 3. Information Extraction-based Multi-Document Summarization

Information Extraction (IE) is a technology that targets the identification of topic-related information in free text and translates it into database entries. Typically, IE systems extract around 10% if a document textual content (cf. (Hobbs and et al.1997)). This represents a compression ratio that qualifies extraction templates for multi-document summarization. This observation was previously employed in the design of the architecture of the SUMMONS multi-document summarization system (Radev and McKeown 1998). In SUMMONS, summarization is viewed as a two-tiered process: (a) *conceptual* and (b) *linguistic* summarization. Conceptual summarization deals with content selection whereas linguistic summarization is concerned with linguistic realization of the content.

To perform conceptual summarization, SUMMONS uses the templates produced by IE to apply a set of *content planning operators* on them for combining the extracted information. These operators, fully detailed in (Radev and McKeown 1998) detect *change of perspective*, *contradiction*, *information addition* or *refinement*. The application of each operator is decided by a set of heuristics, specially crafted for each topic and for each given corpus. The resulting combined templates are then translated into *functional descriptions* (FDs), which are conceptual representations of the template meanings. FDs are used by the linguistic component of SUMMONS that relies on a lexicon and a grammar of English to realize the conceptual representation into a sentence. The linguistic component consists of a lexical chooser, which determines the high-level sentence structure of each sentence and the words that realize each semantic role. SUMMONS incorporates the FUF/SURGE (Elhadad 1993) sentence generator.

In GITEXTER we decided to use IE templates for multi-document summarization in a different way. First we considered not only the populated templates alone, but also the mapping into the text snippets that are the source of their slot fillers. Second, since coreference information is also used to fill slots, we keep pointers to the coreference chains that contain any entity that fills a template slot. Thus for each Template $T_i$ having the slots $TS_i^1$, $TS_i^2$, ..., $TS_i^n$ we keep two additional forms of information: (1) the text snippet $TextS_i^j$ that matched one of the extraction rules, and thus enabled the filling of a slot $TS_i^j$; and (2) all the entities from the text that corefer with the information filling any slot $TextS_i^j$. Figure 9 illustrates a snapshot of populated templates and their mappings. The Figure illustrates some coreference chains as well. Both text snippet information and coreference information is made available by the CICERO IE system.

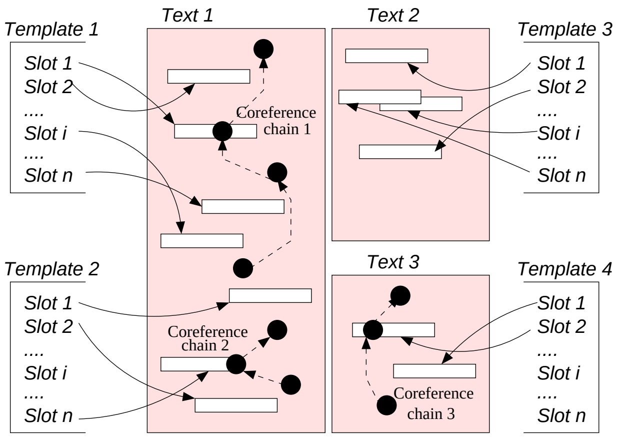


Figure 9: Mappings between extracted templates and text snippets. Whenever a relevant text snippet contains an anaphor, pointers to all other entities with which it corefers are kept in a coreference chain.

To generate multi-document summaries we use two observations: (1) the order in which relevant text snippets appear in the original articles accounts for the coherence of the documents; and (2) to be comprehensible, summaries need to include sentences or sentence fragments that contain the antecedents of each anaphoric expression from relevant text snippets. Since all articles contain information about a given topic, it is very likely that a large percentage of the templates share the same filler for one of the slots. In the case of the “natural disasters” topic, this filler was “ *hurricane Andrew”*. We call this filler the *dominant event* of the collection. Additionally, we are interested in the templates extracting information about other events that may be compared with the dominant event in the collection. Thus templates are classified into four different sets: (a) $Templates_1$ - templates about the dominant event that originate in documents that contain relevant information about related events; (b) $Templates_2$ - other templates about the dominant event; (c) $Templates_3$ - templates about non-dominant events that originate in articles that contain information about the dominant event; and (d) $Templates_4$ - other templates.

To generate a multi-document summary of length $L$ GISTEXTER extracts sentences from the document set in four different increments. The rationale for choosing four increments is based on the four different summary lengths imposed by the DUC evaluations, e.g. 50-word, 100-word, 200-word and 400-word long summaries. Since it is not know apriori how many templates are extracted nor what is the cardinality of each $Templates_i$ set, for each summary increment we perform at least one comparison with the target length $L$ to determine if the resulting summary needs to be reduced or not. The IE-based multi-document summary is produced by the following algorithm:

**Algorithm IE-based MD-Summarization** $(L)$

*Step 1: Select the most representative templates.* To this end, for each template $T_i$ from $Templates_j$, with $1 \leq j \leq 4$, for each slot $TS_i^j$ we count the frequency with which the same filler was used to fill the same slot of any other template. The *importance* of $T_i$ is measured as the sum of all frequency counts of all its slots. This measure generates an order on each of the four sets of templates. Whenever there are ties, we give preference to the template that has the largest number of mapped text snippets traversed by coreference chains. Template $T_0$ is the most important template from $Templates_1$. If $Templates_1$ is null, the same operation is performed on $Templates_2$.

*Step 2: Summary-increment 1.*

Select sentences containing the text snippets mapped from $T_0$ in the order in which they appear in the text from where $T_0$ is selected. If anaphoric expressions occur in any of these sentences, include sentences containing their antecedents in the same order as in the original article.

*if length(summary)* $> L$ generate appositions for dates and locations and drops the corresponding sentences.

*if length(summary)* $> L$ drop coordinated phrases that do not contain any of the mapped text snippets.

*while length(summary)* $> L$ drop the last sentence.

*Step 3: Summary-increment 2.*

For each slot from $T_0$ that has other fillers in some other template from $Templates_1$ or $Templates_2$, add the sentence containing the corresponding mapped text snippet immediately after the sentence mapped by template $T_0$ for the same slot. If anaphoric expressions occur in any of these sentences, include sentences containing their antecedents in the same order as in the original article. Continue this process until either (1) the length of the summary is larger than $L-1$ or until there are no more sentences to be added.

*Step 4: Summary-increment 3.*

Add sentences mapped by the most important template from $Templates_3$. Repeat the process as at Step 2 until length $L$ is reached or no more sentences can be added.

*Step 5: Summary-increment 4.*

Add sentences mapped by the most important template from $Templates_4$. Repeat the process as at Step 2 until length $L$ is reached or no more sentences can be added.

Figure 10 illustrates the inter-leaving of extracted sentences that each summary increment produces in the resulting multi-document summarization.

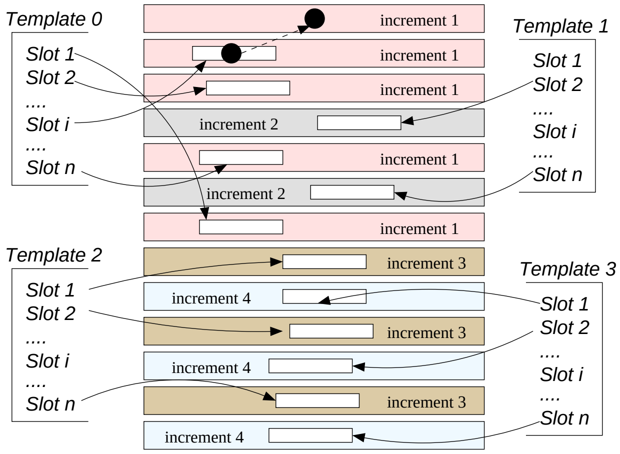


Figure 10: Multi-document summarization produced by four different summary increments.

## 4. Ad-hoc Extraction for Multi-Document Summarization

Whenever the topic of a document collection is not encoded in an IE system, the Algorithm presented in Section 3. cannot be applied. Two main sources of information are missing: (1) the topic template-representation; and (2) the mappings between template slots and text snippets. In (Harabagiu and Maioarano 2000) we have shown that if the template representation of a topic is known, linguistic patterns that identify the mappings of the template slots into text snippets can be acquired automatically. In this paper, we focus on the mechanism of generating the template representation of the topic.

The idea of representing the topic as a frame-like object was first advocated in the late 70's by DeJong (DeJong 1982), who developed a system called FRUMP (Fast Reading Understanding and Memory Program) to skim newspaper stories and extract the main details. The topic representation used in FRUMP is the *sketchy script*, which model a set of pre-defined particular situations, e.g. demonstrations, earthquakes or labor strikes. Since the world contains millions of topics, it is important to be able to generate sketchy script automatically from corpora. In addition some of the current large-scale lexico-semantic knowledge bases may be used to contribute information for the generation of the topic templates. In our methodology, we have employed WordNet (Miller 1995), the lexical database that encodes a majority of the English nouns, verbs, adjectives and adverbs.

### 4.1. Extracting Topical Relations from WordNet

WordNet is both a thesaurus and a dictionary. It is a thesaurus because each word is encoded along with its synonyms in a synonyms set called *synset*, representing a *lexical concept*. WordNet is a dictionary because each synset is defined by a gloss. Moreover, WordNet is a knowledge base because it is organized in 24 noun hierarchies and 512 verb hierarchies. Additionally WordNet encodes three meronym relations (e.g. HAS-PART, HAS-STUFF and HAS-MEMBER) between nouns and two causality relations (e.g. ENTAILMENT and CAUSE-TO) between verbs. However, there are no direct relations between the concepts used in any of the template representation of the topics encoded in the CICERO IE system. Nevertheless we noticed that chains of lexico-semantic relations can be mined from WordNet to account for the connection between any pair of template concepts of known topics. To illustrate how such chains of relations can be mined, we first consider two of the relations already encoded in WordNet and then show how additional relations can be uncovered as lexico-semantic chains between two concepts pertaining to the same topic. We call these lexico-semantic chains *topical relations*.

**The sources of topical relations**

In WordNet, a synset is defined in three ways. First it is defined by the common meaning of the words forming the synset. This definition relies on psycholinguistic principles, based on the human ability to disambiguate a word if several synonyms are presented. Second, the synset is defined by the attributes it inherits from its super-concepts. Third, a glossed definition is provided to each synonym. A GLOSS relation connects a synonym to its definition. We believe that glosses are good sources for topical relations, since they bring forward concepts related to the defined synset. We consider four different ways of using the glosses as sources for topical relations:

1. We extend the GLOSS relation to connect the defined synset not only to a textual definition but to each content word from the gloss, and thus to the synset it represents. For example, the gloss of synset {*bovine spongiform encephalitis, BSE, mad cow disease*} is *( fatal disease of cattle that affects the central nervous system; causes staggering and agitation)*. A GLOSS relation exists between the defined synset and *fatal*, *disease*, *cattle*, *affect*, *central nervous system*, *staggering* and *agitation*.

2. Each concept from a gloss has its own definition, and thus by combining the GLOSS relations, we connect the defined synset to the defining concepts of each concept from its own gloss.

3. The hypernym of a synset has also a gloss, thus a synset can be connected to the concepts from the gloss of its hypernym. Similarly to the IS-A relations, other WordNet lexico-semantic relations can be followed to reach a new synset and have access to the concepts used in its gloss. Such relations may include HAS-MEMBER, HAS-PART or ENTAILS and CAUSE-TO. Lexical relations based on morphological derivations, if available may be used too[4]. Morphological relations include the NOMINALIZATION relations, known to be useful in IE.

4. A synset can be used itself to define other concepts, therefore connections exist between each concept and all concepts it helps define.

[4] WordNet 2 already encodes derivational morphology.

Figure 11 illustrates the four possible sources of topical relations based on two of the WordNet relations, namely GLOSS and IS-A.

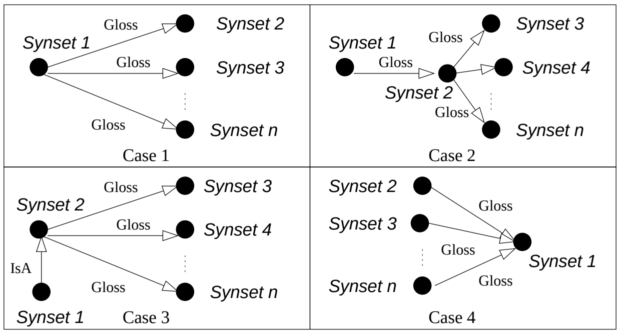


Figure 11: Four sources of topical relations.

**Topical relations as Paths between WordNet Synsets** Two principles guide the uncovering of topical relations. First we believe that redundant connections rule out connection discovered by accident. Therefore, if at least two different paths of WordNet relations can be established between any two synsets, they are likely to be part of the representation of the same topic. Second, the shorter the paths, the stronger their validity. Consequently, we rule out paths of length larger than 4. This entails the fact that each topic may be represented by at least five synsets.

### 4.2. Ad-hoc Templates

A template representation of a topic can be viewed as a list of semantic roles, each role being a slot that is filled by information extracted from text. The topical relations mined from WordNet have the advantage that they bring forward semantically-connected concepts deemed relevant to the topic. However these concepts cannot be mapped directly into a list of slots. First, WordNet was not devised with the IE application in mind - it is a general resource of English lexico-semantic knowledge. Because of this, some concepts relevant to a given topic may not be encoded in WordNet. Second, several WordNet concepts traversed by topical relations may be categorized under the same semantic role. Third, some semantic roles may be encoded in WordNet at a very abstract level, and thus they may never be reached by topical relations. Fourth, some of the semantic roles derived from topical relations may never be filled, since there is no corresponding information in the texts. To address all these issues, we have developed a corpus-based technique for creating ad-hoc lists of semantic roles for the template representation of the collection topic. Our algorithm for ad-hoc template generation was inspired by the empirical approach for conceptual case frame acquisition presented in (Riloff and Schmelzenbach 1998).

**Algorithm Ad-hoc Template Generation**

*Step 1:* Extract all sentences in which one of the concepts traversed by topical relations is present. The concepts from the topical relations are used as a seed lexical items used for the identification of the template slots.

*Step 2:* Identify all Subject-Verb-Object (SVO) +Prepositional attachments syntatic structures in which one of the topical concepts is used. For this purpose, we used the phrasal parser implemented in CICERO as well as all the syntactic variants of the SVO syntactic structures used to implement extraction patterns.

*Step 3:* Apply the IE coreference resolution module and consider all the syntactic SVO structures involving all coreferring expression of any of the nouns used in the syntactic structures discovered at Step 2.

*Step 4:* Combine the extraction dictionaries with WordNet to classify each noun from the structures identified at Step 2 and Step 3.

*Step 5:* Generate the semantic profile of the topic. For this reason we compute three values for each semantic class derived at Step 4: (1) **SFreq**: the number of syntactic structures identified in the collection; (2) **CFreq**: the number of times elements from the same semantic class were identified; and (3) **PRel** the probability that the semantic class identifies a relevant slot of the template. Similarly to the method reported in (Riloff and Schmelzenbach 1998), **PRel** = **CFreq/SFreq**. To select the template slots the following formula is used:

*(* **CFreq** *> F1) or ((***SFreq** *> F2) and (* **PRel** *> P))*

The first test selects roles that because of the semantic categories that are identified with high frequency, under the assumption that this reflect a real association with the topic elaboration in the collection. The second text promotes slots that come from a high percentage of the syntactic structures recognized as containing information relevant to the topic even though their frequency might be low. The values of *F1*, *F2* and *P* vary from one topic to another - we derive them from the requirement that a template should not contain more than 5 slots.

## 5. Evaluation

We participated with GISTEXTER in the DUC-2002 multi-document summarization involving 59 document sets. For each test data set the multi-document summary generated by our system was compared with a gold-standard summary created by humans. For each data set, the author of the gold-standard summary assessed the degree of matching between the model summary and the summaries generated by the systems evaluated in DUC-2002.

Each of these measures were scored on a scale between 0 and 4.

To compute the quantitative measures of overlap between the system-generated summaries and the gold-standard summary, the human-created summary was segmented by hand by assessors into *model units* (MUs), which are informational units that should express one self-contained fact in the ideal case. MUs are sometimes sentence clauses, sometimes entire clauses. In contrast, the summaries generated by the summarization systems were automatically segmented into *peer units* (PUs) - which are always sentences. Figure 12 lists the results obtained for the single-document summarization evaluations. By ranking according to the mean coverage of PUs into MUs and the respective median coverage, GISTEXTER, labeled as system 19, was ranked as the first system. For mean-length adjusted coverage it was ranked on the second place whereas for median length-adjusted coverage it was ranked on the third place.

SINGLE-DOC

| System | Score for quality questions with non-0 answers | Count of quality questions with non-0 answers | Mean score for quality questions with non-0 answers | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 | Q11 | Q12 | Fraction of unmarked peer units at least related to the model's subject | Number of peer units | Number of marked peer units | Number of unmarked peer units | Number of model units | Mean coverage | Median coverage | Sample standard deviation of coverage scores | Mean length-adjusted coverage | Median length-adjusted coverage | Sample standard deviation of adjusted coverage scores |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **15** | 0.54 | 0.99 | 0.55 | 0.1 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.1 | 0.3 | 0.6 | 4.1 | 2.8 | 1.3 | 9.4 | 0.33 | 0.20 | 0.40 | 0.22 | 0.13 | 0.27 |
| **16** | 0.93 | 1.44 | 0.64 | 0.5 | 0.1 | 0.0 | 0.2 | 0.2 | 0.1 | 0.1 | 0.2 | 0.0 | 0.0 | 0.1 | 0.2 | 0.3 | 2.7 | 2.1 | 0.6 | 9.4 | 0.30 | 0.18 | 0.38 | 0.20 | 0.12 | 0.26 |
| **17** | 0.31 | 0.76 | 0.41 | 0.06 | 0.03 | 0 | 0.12 | 0.05 | 0.02 | 0.06 | 0.22 | 0.014 | 0.02 | 0.04 | 0.13 | 0.23 | 1.31 | 0.76 | 0.56 | 9.44 | 0.08 | 0.01 | 0.15 | 0.26 | 0.22 | 0.1 |
| **18** | 0.56 | 1 | 0.57 | 0.14 | 0.04 | 0.01 | 0.13 | 0.08 | 0.02 | 0.12 | 0.23 | 0.01 | 0.03 | 0.04 | 0.19 | 0.37 | 2.8 | 2.15 | 0.65 | 9.43 | 0.32 | 0.19 | 0.4 | 0.22 | 0.13 | 0.26 |
| **19** | 0.31 | 0.7 | 0.45 | 0.13 | 0.04 | 0 | 0.08 | 0.03 | 0.01 | 0.05 | 0.13 | 0.017 | 0.01 | 0.04 | 0.19 | 0.49 | 3.88 | 3.03 | 0.85 | 9.41 | 0.39 | 0.29 | 0.43 | 0.27 | 0.2 | 0.29 |
| **21** | 0.5 | 0.88 | 0.56 | 0.1 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.1 | 0.2 | 0.5 | 3.6 | 2.7 | 0.9 | 9.4 | 0.37 | 0.25 | 0.42 | 0.25 | 0.17 | 0.28 |
| **23** | 0.25 | 0.58 | 0.42 | 0.2 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.1 | 0.4 | 3.2 | 2.5 | 0.7 | 9.4 | 0.34 | 0.18 | 0.42 | 0.23 | 0.13 | 0.28 |
| **25** | 4.1 | 3.2 | 1.28 | 1.95 | 0.13 | 0.08 | 0.69 | 0.22 | 0.07 | 0.13 | 0.37 | 0.017 | 0.04 | 0.11 | 0.49 | 0.61 | 4.24 | 2.44 | 1.81 | 9.39 | 0.29 | 0.16 | 0.38 | 0.2 | 0.11 | 0.25 |
| **27** | 0.56 | 1.01 | 0.55 | 0.14 | 0.03 | 0.02 | 0.12 | 0.05 | 0.02 | 0.11 | 0.19 | 0.01 | 0.04 | 0.07 | 0.26 | 0.52 | 4.43 | 3.14 | 1.29 | 9.49 | 0.38 | 0.27 | 0.43 | 0.26 | 0.19 | 0.29 |
| **28** | 0.54 | 1.01 | 0.54 | 0.3 | 0.1 | 0.0 | 0.1 | 0.1 | 0.0 | 0.1 | 0.1 | 0.0 | 0.0 | 0.1 | 0.2 | 0.4 | 3.4 | 2.7 | 0.7 | 9.4 | 0.38 | 0.27 | 0.43 | 0.25 | 0.18 | 0.29 |
| **29** | 0.8 | 1.21 | 0.66 | 0.2 | 0.0 | 0.0 | 0.2 | 0.1 | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.1 | 0.3 | 0.6 | 4.1 | 2.8 | 1.3 | 9.4 | 0.36 | 0.24 | 0.42 | 0.24 | 0.16 | 0.28 |
| **30** | 2.75 | 2.64 | 1.04 | 1.1 | 0.2 | 0.1 | 0.7 | 0.2 | 0.1 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 1.0 | 0.7 | 0.3 | 9.4 | 0.06 | 0.00 | 0.14 | 0.30 | 0.26 | 0.09 |
| **31** | 0.78 | 1.15 | 0.68 | 0.2 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.1 | 0.3 | 0.6 | 4.3 | 2.9 | 1.4 | 9.5 | 0.36 | 0.26 | 0.40 | 0.24 | 0.17 | 0.27 |
| **Ranks** | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **15** | 5 | 5 | 5 | 2 | 4 | 9 | 9 | 2 | 1 | 7 | 5 | 13 | 12 | 7 | 9 | | | | | | 8 | 7 | | 10 | 9 | |
| **16** | 11 | 11 | 9 | 11 | 11 | 6 | 11 | 13 | 11 | 6 | 11 | 3 | 9 | 9 | 8 | | | | | | 10 | 9 | | 12 | 12 | |
| **17** | 2 | 3 | 1 | 1 | 7 | 1 | 5 | 5 | 8 | 5 | 10 | 9 | 7 | 2 | 3 | | | | | | 12 | 12 | | 3 | 2 | |
| **18** | 8 | 6 | 8 | 4 | 8 | 5 | 6 | 8 | 7 | 10 | 12 | 4 | 8 | 3 | 6 | | | | | | 9 | 8 | | 11 | 10 | |
| **19** | 3 | 2 | 3 | 3 | 9 | 2 | 2 | 1 | 5 | 2 | 3 | 10 | 4 | 3 | 5 | | | | | | 1 | 1 | | 2 | 3 | |
| **21** | 4 | 4 | 7 | 4 | 1 | 3 | 3 | 7 | 1 | 9 | 4 | 4 | 5 | 6 | 7 | | | | | | 4 | 5 | | 6 | 7 | |
| **23** | 1 | 1 | 2 | 7 | 3 | 4 | 1 | 6 | 4 | 3 | 1 | 2 | 2 | 5 | 2 | | | | | | 7 | 10 | | 9 | 11 | |
| **25** | 13 | 13 | 13 | 13 | 12 | 12 | 12 | 12 | 12 | 12 | 13 | 11 | 11 | 13 | 13 | | | | | | 11 | 11 | | 13 | 13 | |
| **27** | 7 | 7 | 6 | 6 | 4 | 11 | 4 | 4 | 9 | 8 | 6 | 4 | 12 | 10 | 10 | | | | | | 2 | 2 | | 4 | 4 | |
| **28** | 6 | 7 | 4 | 10 | 10 | 6 | 8 | 10 | 10 | 4 | 2 | 7 | 3 | 11 | 4 | | | | | | 3 | 3 | | 5 | 5 | |
| **29** | 10 | 10 | 10 | 9 | 4 | 6 | 10 | 9 | 5 | 11 | 7 | 12 | 6 | 12 | 11 | | | | | | 5 | 6 | | 7 | 8 | |
| **30** | 12 | 12 | 12 | 12 | 13 | 13 | 13 | 11 | 13 | 1 | 9 | 1 | 1 | 1 | 1 | | | | | | 13 | 13 | | 1 | 1 | |
| **31** | 9 | 9 | 11 | 8 | 2 | 9 | 6 | 3 | 1 | 13 | 8 | 7 | 10 | 7 | 12 | | | | | | 6 | 4 | | 8 | 6 | |

Figure 12: Results of the single-document summarization evaluations in DUC-2002

Figure 12 lists also the results of the evaluations with respect to the accuracy with which the summaries responded the twelve questions listed in Figure 13.

*Q1: About how many gross capitalization errors are there?*

*Q2: About how many sentences have incorrect word order?*

*Q3: About how many times does the subject fail to agree in number with the verb?*

*Q4: About how many of the sentences are missing important components (e.g. the subject, main verb, direct object, modifier) −causing the sentence to be ungrammatical, unclear or misleading?*

*Q5: About how many times are unreleted fragments joined into one sentence?*

*Q6: About how many times are articles (a, an, the) missing or used incorrectly?*

*Q7: About how many pronouns are there whose antecedents are incorrect, unclear, missing or come only later?*

*Q8: About how many nouns is it impossible to deterine clearly who or what they refer to?*

*Q9: About how many times should a noun or noun phrase have been replaced with a pronoun?*

*Q10: About how many dangling conjunctions are there ("and", "however" ...)?*

*Q11: About how many instances of repeated information are there?*

*Q12: About how many sentences strike you as in the wrong place because they indicate a strange time sequence, suggest a wrong cause−effect relationship, or just don't fit in topically with neighboring sentences?*

Figure 13: Qualitative questions used to evaluate summaries in DUC-2002.

Figure 15 lists similar results for the multi-document summarization evaluations for abstracts. By ranking according to the mean coverage of PUs into MUs and the respective median coverage, GISTEXTER was ranked as the first system. It was also ranked as the first system for mean-length adjusted coverage and for median length-adjusted coverage.

| **System** | Precision | Recall | Rank P | Rank R |
|---|---|---|---|---|
| **16** | 0.1219 | 0.078909 | 9 | 9 |
| **19** | 0.206647 | 0.207082 | 3 | 1 |
| **20** | 0.148241 | 0.152151 | 5 | 5 |
| **21** | 0.249052 | 0.206362 | 1 | 2 |
| **22** | 0 | 0 | 10 | 10 |
| **24** | 0.221155 | 0.182388 | 2 | 3 |
| **25** | 0.130039 | 0.10453 | 6 | 8 |
| **28** | 0.181207 | 0.158358 | 4 | 4 |
| **29** | 0.12797 | 0.120487 | 7 | 6 |
| **31** | 0.126711 | 0.107595 | 8 | 7 |

Figure 14: Results of the multi-document summarization evaluations for extracts in DUC-2002.

For multi-document summaries, we considered also the *Precision* and *Recall* measures. Precision is calculated as the number of PUs matching some MU divided by the number of PUs in the peer summary, considering all summaries automatically generated for the same collection. Figure 14 lists the precision and recall results for all the systems that participated in DUC-2002. Our system was ranked on the second place for precision and first place for recall measures. As reported in (McKeown et al.2001), this estimate of the precision is conservative, since the number of PUs that are considered correct can be increased by considering information about the PUs not assigned to MUs.

## 6. Conclusions

In this paper we have shown that multi-document summarization of good quality can be obtained if extraction

MULTI-DOC

| System | Score for quality questions with non-0 answers | Count of quality questions with non-0 answers | Mean score for quality questions with non-0 answers | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 | Q11 | Q12 | Fraction of unmarked peer units at least related to the model's subject | Number of peer units | Number of marked peer units | Number of unmarked peer units | Number of model units | Mean coverage | Median coverage | Sample standard deviation of coverage scores | Mean length-adjusted coverage | Median length-adjusted coverage | Sample standard deviation of adjusted coverage scores |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **16** | 1.11 | 1.7 | 0.65 | 0.6 | 0.1 | 0.0 | 0.3 | 0.4 | 0.1 | 0.1 | 0.3 | 0.0 | 0.0 | 0.2 | 0.5 | 0.3 | 2.1 | 1.5 | 0.6 | 8.6 | 0.13 | 0.05 | 0.17 | 0.09 | 0.04 | 0.11 |
| **19** | 0.52 | 1 | 0.52 | 0.32 | 0.07 | 0.02 | 0.14 | 0.08 | 0.11 | 0.07 | 0.17 | 0.045 | 0 | 0.15 | 0.32 | 0.39 | 4.19 | 2.98 | 1.21 | 8.62 | 0.28 | 0.19 | 0.25 | 0.19 | 0.13 | 0.17 |
| **20** | 1.08 | 1.54 | 0.7 | 0.31 | 0.05 | 0.04 | 0.19 | 0.07 | 0.03 | 0.16 | 0.36 | 0.028 | 0.08 | 0.16 | 0.77 | 0.4 | 4.05 | 2.3 | 1.75 | 8.62 | 0.15 | 0.07 | 0.19 | 0.11 | 0.05 | 0.13 |
| **24** | 1.71 | 2.07 | 0.83 | 0.23 | 0.19 | 0.07 | 0.31 | 0.13 | 0.08 | 0.07 | 0.25 | 0.034 | 0.06 | 0.2 | 0.54 | 0.4 | 3.94 | 2.75 | 1.19 | 11 | 0.18 | 0.06 | 0.27 | 0.13 | 0.05 | 0.18 |
| **25** | 2.66 | 2.81 | 0.95 | 1.86 | 0.21 | 0.07 | 0.78 | 0.31 | 0.14 | 0.1 | 0.37 | 0.017 | 0.1 | 0.18 | 0.65 | 0.45 | 3.7 | 1.96 | 1.75 | 8.62 | 0.13 | 0.05 | 0.17 | 0.09 | 0.04 | 0.11 |
| **26** | 1.51 | 2.03 | 0.74 | 0.5 | 0.1 | 0.0 | 0.5 | 0.3 | 0.2 | 0.1 | 0.3 | 0.1 | 0.0 | 0.3 | 0.6 | 0.5 | 4.4 | 2.9 | 1.6 | 8.6 | 0.22 | 0.12 | 0.23 | 0.15 | 0.08 | 0.16 |
| **28** | 1.48 | 1.69 | 0.87 | 0.4 | 0.1 | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.2 | 0.6 | 0.3 | 3.6 | 2.8 | 0.8 | 11.0 | 0.22 | 0.09 | 0.30 | 0.15 | 0.06 | 0.20 |
| **29** | 1.26 | 1.77 | 0.71 | 0.2 | 0.0 | 0.0 | 0.3 | 0.1 | 0.0 | 0.4 | 0.4 | 0.0 | 0.1 | 0.1 | 0.8 | 0.4 | 3.9 | 2.0 | 1.9 | 8.6 | 0.14 | 0.06 | 0.19 | 0.10 | 0.05 | 0.13 |
| **Ranks** | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **16** | 3 | 4 | 2 | 7 | 5 | 3 | 5 | 8 | 4 | 6 | 4 | 5 | 3 | 3 | 2 | | | | | | 7 | 7 | | 8 | 8 | |
| **19** | 1 | 1 | 1 | 4 | 3 | 3 | 1 | 2 | 6 | 3 | 2 | 6 | 1 | 2 | 1 | | | | | | 1 | 1 | | 1 | 1 | |
| **20** | 2 | 2 | 3 | 3 | 2 | 6 | 3 | 1 | 1 | 7 | 6 | 3 | 7 | 3 | 7 | | | | | | 5 | 4 | | 5 | 4 | |
| **24** | 7 | 7 | 6 | 1 | 7 | 7 | 6 | 3 | 5 | 3 | 3 | 4 | 5 | 6 | 3 | | | | | | 4 | 6 | | 4 | 5 | |
| **25** | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 7 | 7 | 5 | 7 | 1 | 8 | 5 | 6 | | | | | | 8 | 8 | | 7 | 7 | |
| **26** | 6 | 6 | 5 | 6 | 5 | 2 | 7 | 6 | 8 | 2 | 5 | 7 | 3 | 8 | 5 | | | | | | 2 | 2 | | 2 | 2 | |
| **28** | 5 | 3 | 7 | 5 | 4 | 1 | 1 | 5 | 3 | 1 | 1 | 8 | 1 | 7 | 4 | | | | | | 3 | 3 | | 3 | 3 | |
| **29** | 4 | 5 | 4 | 2 | 1 | 5 | 4 | 4 | 1 | 8 | 8 | 2 | 6 | 1 | 8 | | | | | | 6 | 5 | | 6 | 6 | |

Figure 15: Results of the multi-document summarization evaluations for abstracts in DUC-2002.

templates populated by a high performance IE systems are available. We have presented an IE-based multi-document summarization procedure that incrementally adds information to create summaries of variable size. The decision of using incremental additions of sentences from multiple documents based on their mapping from the template slots produced very good results for coherence and organization in the DUC-2002 evaluations.

## 7. References

R. Barzilay, N. Elhedad and K. McKeown. Sentence ordering in multidocument summarization. In *Proceedings of Conference on Human Language Technology (HLT-2001)*, 2001.

G. DeJong. An overview of the FRUMP system. In *Stategies for natural language processing*, W. Lehnert and M. Ringle Eds., pages 149-176, Lawrence Erlbaum Associates, 1982.

M. Elhadad. Using argumentation to control lexical choice: A unification-based implementation. PhD Thesis, Computer Science Department, Columbia University, 1993.

S. Harabagiu and S. Maiorano. Acquisition of Linguistic Patterns for Knowledge-Based Information Extraction. In *Proceedings of Second International Conference on Language Resources and Evaluation (LREC-2000)*, Athens, Greece, 2000.

D. Appelt, J. Hobbs, J. Bear, D. Israel, M. Kameyama, M. Stickel and M. Tyson. FASTUS: A Cascaded Finite-State Transducer for Extracting Information from Natural-Language Text. In *Finite State Language Processing*, edited by Emmanuel Roche and Yves Schabes, MIT Press, 1997.

C.Y. Lin and E. Hovy. The Automated Acquisition of Topic Signatures for Text Summarization. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000)*, Saarbrcken, Germany, July 31- August 4, 2000.

I. Mani and M. Maybury, eds. Advances in Automatic Text Summarization. MIT Press, 1999.

K. McKeown, R. Barzilay, D. Evans, V. Hatzivassiloglou, M.Y. Kan, B. Schiffman and S. Teufel. Columbia Multi-Document Summarization: Approach and Evaluation. In *Workshop Notes for the DUC-2001 Summarization*, pages 43–64, September 2001.

George A. Miller. WordNet: a lexical database for English. In *Communications of the ACM*, Vol.38, No.11:39–41, 1995.

D. Radev and K. McKeown. Generating natural language summaries from multiple on-line sources. *Computational Linguistics*, 24(3):469-500, September 1998.

E. Riloff and M. Schmelzenbach. An Empirical Approach to Conceptual Case Frame Acquisition. In *Proceedings of the Sixteenth Workshop on Very Large Corpora*, 1998.

E. Riloff and R. Jones. Learning Dictionaries for Information Extraction by Multi-Level Bootstrapping. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*, pp. 474-479, 1999.

M. Surdeanu and S. Harabagiu. Infrastructure for Open-Domain Information Extraction. In *Proceedings of Conference on Human Language Technology (HLT-2002)*, 2002.