LITE-GISTEXTER: Generating Short Summaries with Minimal Resources

Valeriu Finley Lăcătuşu, Paul Parker* and Sanda M. Harabagiu*,

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

Abstract

This paper presents the techniques implemented in LITE-GISTEXTER for producing short summaries from both single and multiple documents. These techniques use minimal lexico-semantic resources and produce summaries of acceptable quality. The results of LITE-GISTEXTER in the DUC-2003 were surprising, especially for the Task of answering questions with multi-document summaries.

1. Introduction

Short summaries of documents allow the user to rapidly scan through text collections, in search of a particular piece of information. When summaries contain at most 10 words, they act like headlines, conveying the most important information in as few words as possible. Such headlines may be indexed, enabling a different kind of retrieval than search engines currently provide. Moreover, multi-document summaries can be used to rank the relevance of document clusters as well. Unlike current retrieval systems that typically index documents by content words only, summaries focused on specific events or based on given viewpoints enable the development of information searching tools that personalize the user intentions.

The ultimate way of searching for information is by simply posing a question in natural language and obtaining an answer containing the desired knowledge. One of the most intriguing applications of multi-document summaries is provided by textual Question/Answering(Q/A). Given a natural language question, often its answer is derived from multiple documents, especially in the case of complex questions. Trivia-like questions, similar to those evaluated in the TREC tracks, can be answered if two main resources are available: (1) Named Entity recognizers and (2) lexico-semantic knowledge bases that allow us to model the semantics of the relations between a question and its answer. In the case of Q/A systems, one of the main difficulties stems for the need to identify in large text collections candidate answers with acceptable precision. The other challenge is to extract the correct answer from a list of candidate answers. In DUC-2003 the solution to the first problem was given, as a set of relevant documents was provided along with their sentences containing the information deemed useful to the answer. Instead of extracting the answer, we had to generate a summary from the relevant sentences, as a possible answer to the question. This task offered an interesting alternative to the paradigm currently used in Q/A, in which different forms of unifications between a question and its answers are modeled.

In DUC-2003 four different kinds of short summaries were evaluated. Task 1 targeted very short summaries of single documents. As a minimum, this requires a mechanism of selecting useful sentences and a compression technique that reduces a sentence to a maximum of 10 words. The simplest way of selecting useful sentences is by always choosing the first sentence in the document. Sentence compression may be achieved by considering syntactic information, thus a parser is required as a minimal resource. Task 2 targeted multi-document short summaries focused on a seminal event. For this task, sentence selection is more complex, therefore we relied on WordNet, a vast lexico-semantic resource that enabled us to implement similarity measures between each sentence and the event description. Consequently, sentence selection was based on the similarity ranking. Task 3 targeted multi-document summaries based on a viewpoint. WordNet again was used to select sentences displaying similarity to the viewpoint description. Task 4 from DUC-4 attracted us most. It required to produce multi-document summaries in response to a question. The questions were very different from the questions evaluated in the TREC Q/A tracks. Most of the time there were no identifiable expected answer type, thus Q/A systems could no perform at their best on such questions. However, we found that the assumption that semantic information plays an important role in Q/A could once again be verified, even when minimal resources are involved. By training a classifier to select sentences for the summary that answer the question, we found that named entity information along with semantic relations encoded in WordNet provide the sufficient semantic resources for this task. Figure 1 shows the architecture of LITE-GISTEXTER and the readily available resources it employs.

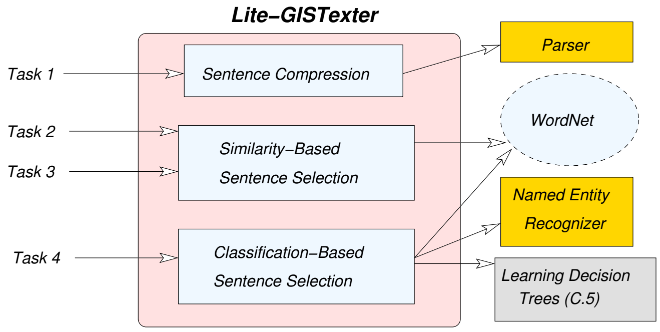

Figure 1: LITE-GISTEXTER in DUC-2003

2. Generating very short summaries

A very short document summary is generated with the purpose of allowing the user to decide with minimal effort

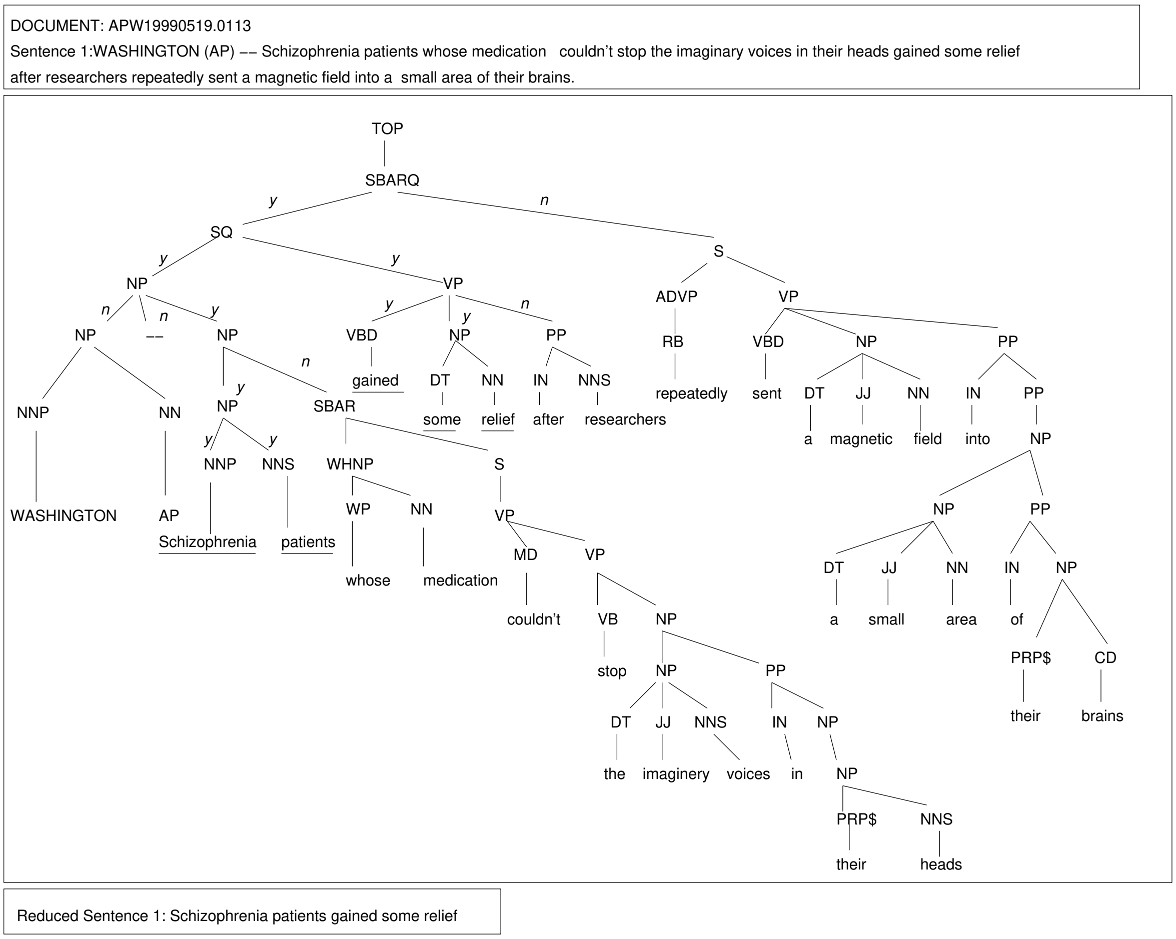

Figure 2: Sentence compression through syntactic information

whether the document is useful to a given task or not. The effort may be measured in many ways. A simple measure is dictated by the number of words the user needs to read and interpret. Given this quantification of effort, methods of generating very short summaries become relevant. In DUC-2003, in Task 1 the length of very short summaries was constrained to a maximum of 10 words. Such very summaries can also be used as documents headlines.

Two sets of documents were given: (1) a set of 30 TREC document clusters related by subject of interest; and (2) a set of 30 TDT document clusters related by topics or events or timespans. Each cluster contained an average of 10 documents. Documents from both collections originated from the following news sources:

- AP newswire, 1998-2000;
- New York Times newswire, 1998-2000; and
- Xinhua News Agency (English version), 1996-2000

To create a very short single-document summary we fist need to identify the most useful information in the document and extract it in a maximum of ten words. Out assumption was that generally, textual information is cohesive, and thus the most useful information is likely to be found in a single sentence. Moreover, from previous DUC evaluations, we were aware that the average length of document sentences surpasses 20 words, therefore it made sense to create the very short summary from a single sentence.

Many single-document summarization techniques rely on the assumption that the most useful information is located in the first document paraphraph (cf. (Mani and Maybury 1999), (Lin and Hovy 2000)). We took this assumption further, and considered that the sentence containing the most useful information is the first sentence of the first paragraph, or the first sentence of the document. This very simplistic sentence selection mechanism can be improved by modeling sentence usefulness with statistical methods or with NLP methods that rely on lexico-semantic information. In the absence of such techniques, selecting the first sentence on the document may be considered as a baseline method.

More than often, the first sentence is longer than 10 words. In the rare case when its length is smaller than the limit of 10 words, the first sentence in its entirety becomes the short summary of the document. For the more frequent case, we had to generate a sentence compression mechanism that reduces the sentence to a 10-word summary. In the process we wanted the compression technique to preserve the information considered useful. Our way of modeling the useful information is based on the notion of "trigger word". This concept was introduced in (Kraaij et al.2002) as representing the most frequent word in a document cluster. Kraaij and his colleagues used the "trigger word" for

Sentence	Trigger Word
Schizophrenia patients whose medication couldn't stop the imaginary voices in their heads gained some relief after researchers repeatedly sent a magnetic field into a small area of their brains.	schizophrenia
Water levels at the middle and lower reaches of the Yangtze River, China's longest, started to recede today after the largest flood peak this year passed Wuhan, capital of central China's Hubei Province.	Yangtze
Charles Schulz, the creator of ''Peanuts,'' the tender and sage comic strip starring Charlie Brown and Snoopy that was read by 355 million people around the world, died in his sleep on Saturday night at his home in Santa Rosa, Calif., just hours before his last cartoon ran the Sunday newspapers.	Schultz
Auctioning on the Internet is one of the fastest−growing forms of electronic commerce, on a pace to reach $40 billion a year, according to some experts.	auctioning

Figure 3: Examples of trigger words derived from the first sentence of a document

generating headline-like multi-document summaries.

For the DUC-2003 Task 1 we redefined the notion of "trigger word" as the word from the first sentence that has the highest frequency in the document. We have also considered a list of 573 "common words". Only content words different from the "common words" were considered possible "trigger words". Figure 3 illustrates some of the trigger words considered from the first sentences of several of the summarized documents. We assigned trigger words an important role in sentence reduction, as we allowed them to always appear in the summary. All the other words from the first document sentence may be eliminated. The decision of keeping or dropping a word from the first sentence is based on information derived from the full parse tree. The parse was obtained by a probabilistic parser similar to the Collins parser (Collins 1996).

As first reported in (Jing 2000), the sentence reduction process can be considered as a series of decisions made along keeping or dropping an edge of the parse tree. If an edge is dropped, the subtree at the address of the edge is completely discarded. Figure 2 shows the parse tree obtained for the first sentence from document APW19990519.0113 as well as the reduction decisions that produced the compressed sentence. The reduction decisions were made based only on 26 reduction rules. The rules were expressed as search patterns, in which three forms of search on the parse tree were allowed: (1) breadth-first search (BFS); (2) in-order search (IOS); and (3) depth-first search (DFS). Each rule corresponds to a path between the top of the parse tree and a constituent that is kept in the reduced sentence. The format of the decision rules consists of two ordered lists: (1) a list of triplets $\{T_i\}$=$\{<i,S_i,N_i>\}$, where i represents the index in the path, S_i stands for the type of search performed between the current node and the successive node and N_i is the label of the node, either a terminal or a non-terminal; (2) a list of binary decisions $\{D_i\}$ such that $D_i = 1$ indicates that the edge between N_i and N_{i+1} is marked with y whereas $D_i = 0$ will mark it as n. For example, the decision rule that retained the noun phrase *"Schizophrenia patients"* from the parse illustrated in Figure 2 was: [{< 1, BFS, TOP>, < 2, DFS, SQ >, < 3, BFS, NP > < 4, DFS, NP >, < 5, DFS, NP >}, { 1 1 0 1 1}].

Figure 4 shows the evalaution of the usefulness of the information retained in the very short summaries produced by our method. The results indicate that perhaps we need to experiment with more informed methods of sentence selection.

TASK 1 **USEFULNESS**	7	7
	8	1
	9	9
	10	13
	13	10
	15	12
	17	2
	18	3
	21	5
	22	**8**
	24	11
	25	6
	26	4

Figure 4: Results of usefulness.

3. Short summaries focused by events

Creating a multiple-document summary focused on a given event, described as a TDT topic was the focus of Task 2 in DUC-2003. The goal was to generate for each of the 30 TDT clusters[1] a short summary of at most 100 words. Figure 5 shows an example of a topic description and its seminal event description as a frame-like template with the slots filled by information pertinent to the seminal event. As it can be seen, the template needs not be used for extracting any other similar events, since only one seminal event is sought. Consequently we found that typical Information Extraction (IE) techniques are not useful for generating short summaries focused by a single event, since such techniques are more useful for identifying information relevant for a class of events rather than a unique event.

We decided to focus instead on the topic description as a more detailed source of information for the event of interest. With the goal of building a short summary of at most 100 words that covers the event of interest, we developed the following three step procedure:

<u>*Step 1:*</u> For each sentence in each document
  Measure the topic-similarity of each content word
    to the topic description;
<u>*Step 2:*</u> For each sentence in each document
  Generate the topic-similarity rank as the sum
    of the topic-similarities of all its content words;
<u>*Step 3:*</u> Select the sentences having the highest
  topic-similarities ranks.

[1]The same TDT clusters were also used in Task 1 of DUC-2003

Topic:31022. Swedish Dance Hall Fire

Seminal Event:

WHAT: Fire erupted in a Swedish dance hall during a Halloween party, where 400 people were in attendance
WHO: Mostly immigrant teenagers between the ages of 12–20
WHERE: Goteberg, Sweden
WHEN: October 29, 1998

Topic Explication

Fire erupted in a Swedish dance hall during a Halloween party where 400 young people were in attendance. The maximum capacity of the building was 150. Authorities blamed the explosive speed of the blaze and the overcrowding of the building for the 62 deaths and 162 injuries that occurred. Witnesses at the event blame the slow and unhelpful response of the fire trucks and rescue squads.
On topic: stories discussing any aspect of this fire are on topic.
For example, stories on casualities and injuries, response of the emergency teams, and interviews with authorities and witnesses.

Topic Research

Mandarin Resources

Rule of Interpretation: 5: Accidents

Seed Story: APW19981102.0190

Figure 5: Example of a topic and its seminal event.

The topic-similarity of a word w_i to a topic T is measured as the sum of the word-to-word similarities between word w_i to each of the content words w_j from the topic description $D(T)$. The word-to-word similarity score is a value between 0 and 1 computed as follows:

- if the words w_i and w_j are identical, the score is 1;
- if the words w_i and w_j belong to the same WordNet synset, the score is 0.95;
- if the words w_i and w_j are one in the other's WordNet gloss, the score is 0.3;
- if the words w_i and w_j belong to two distinct WordNet synsets S_a and S_b and there is a sequence of n IS-A relations between S_a and S_b, the score is equal to $1/n$;
- if the words w_i and w_j belong to two distinct WordNet synsets S_a and S_b and there is a sequence of n IS-PART relations between S_a and S_b, the score is equal to $1/2n$.

Figure 6 shows the multi-document short summary produced by our technique.

A fire turned a dance hall jammed with teen–age Halloween revelers into a deathtrap, killing 60 people and injuring 162 others in Sweden's second–largest city.

The dead, as young as 12 and none older than 20, were among an estimated 400 people who were at a disco dance in a rented second–floor hall when an explosive fire broke out shortly before midnight Thursday.

Figure 6: Example of an event-focused summary.

4. Short summaries focused by viewpoints

Another way of generating multi-document summaries is possible when a given viewpoint is considered. Task 3 from DUC-2003 focused on this form of summarization. For each of the 30 TREC clusters considered as data[2] a viewpoint description was given in natural language. Figure 7 illustrates the viewpoint given for one of the clusters of documents. The viewpoint was defined as describing the important facets of each cluster as decided by a human assessor. Each facet was represented in at least all but one of the documents in the cluster. However, as seen in Figure 7, the facets are not marked up, but they could be loosely associated with each clause in the viewpoint description.

Forty years after poor parenting was thought to be the cause of schizophrenia, researchers are working in many diverse areas to refine the causes and treatments of this disease and enable early diagnosis.

Figure 7: Example of viewpoint.

Viewpoint-focused multi-document summaries can be created by selecting sentences relevant to each facet of the viewpoint or relevant to the entire viewpoint. If the latter choice is made, a similarity metric using the same scoring mechanism as the one we employed for Task 2 can be used. Figure 8 shows the result of this technique when the viewpoint from Figure 7 is used.

Schizophrenia patients whose medication couldn't stop the imaginary voices in their heads gained some relief after researchers repeatedly sent a magnetic field into a small area of their brains.
Dr. Thomas H. McGlashan, the director of the Yale study, said he decided to begin his prevention trial when he became convinced, after many years of working with patients suffering from schizophrenia, that "80 percent if not more of the damage is done before the disorder appears."

Figure 8: Example of a viewpoint-based summary.

5. Short summaries that answer a question

Summaries may represent an interesting way of answering a question, especially when the question is not a trivia-like question, asking "Who is the queen of England?" or "How did Adolf Hitler die?". From all the tasks evaluated in DUC-2003, this task, namely Task 4, attracted us the most. The data used in this task consists of 30 TREC Novelty track clusters. For each question the documents were ranked according to the relevance to the question. Additionally, a set of sentences that contained relevant information was identified in each document. Given the ranked documents as a list D_1, D_2, ..., D_{30}, the relevant sentences are also organized as a list indexed along two dimensions: (1) the document containing the sentence; and (2) the order of the sentence in the document. The list of relevant sentences may be represented as: R=$S_1^{D_1}$, $S_2^{D_1}$,...,$S_k^{D_1}$, $S_1^{D_2}$, $S_2^{D_2}$,...,$S_q^{D_2}$,$S_1^{D_3}$,...,$S_j^{D_{30}}$.

[2]The same TREC clusters were also used in Task 1 of DUC-2003.

The novel sentences are a subset of the relevant sentences. A sentence from document D_i is novel if any information contained in was not mentioned in any of the sentences preceding it in the list of sentences R. Clearly, the list of novel sentences N follows the same dual-index ranking as the list R. In the generation of short summaries that respond to questions we used only sentences from the N list.

As the answer has to be related semantically to the question, we have developed a machine learning mechanism based on decision trees for selecting which novel sentences should be included in the answer. The selection grouped sentences in three classes: H - sentences for which we have high confidence they belong to the answer; M - sentences for which we have marginal confidence in the decision of including them in the answer; and L - sentences for which we have low confidence in including them in the answer.

For each topic, we tagged the novel sentences with the values of three sets of features:

1. *Lexico-semantic features* based on information available from WordNet. For each sentence S from N, we generate all pairs of content words (w_i, w_j), in which w_i belongs to the question whereas w_j belongs to S. Both words w_i and w_j are selected in a left-to-right order. Each such pair of words (w_i, w_j) allows us to generate values for 18 lexico-semantic features.
   - Whenever we find in WordNet a sequence of IS-A relations that connects one of the semantic senses of w_i with one of the semantic senses of w_j, we tag sentence S with a value representing the length of the relation sequence. We considered a total of such seven tags, representing the values of features H_1, H_2, ..., H_7 that encode the lengths of the first seven pairs of words related by IS-A (hypernym) links.
   - Whenever we find in WordNet a sequence of IS-PART relations that connects one of the semantic senses of w_i with one of the semantic senses of w_j, we tag sentence S with a value representing the length of the relation sequence. We considered a total of such seven tags, representing the values of features M_1, M_2, ..., M_7 encoding the lengths of the first seven pairs of words related by IS-PART (meronym) links.
   - Feature *In-Synset* is set to 1 whenever words w_i and w_j belong to the same WordNet synset.
   - Feature *In-Gloss* is set to 1 whenever word w_i belongs to the defining WordNet gloss of w_j.
   - Feature *Hypernym-Aggregate (FA)* is computed as $FA = \sum_{i=1}^{7} \frac{1}{H_i}$
2. *Semantic class features* annotated by a Named Entity tagger. The role of semantic information provided by Named Entities in Question Answering was from the first TREC evaluations surprisingly important (cf. (Srihari et al.2000), (Radev et al.1999)). In consequence we used 9 features that encode information related to the recognition of named entities.
   - Feature *NE-Total* counts the number of named entities recognized in each sentence S from the list of novel sentences N.
   - Feature *NE-Location* counts the number of locations recognized in each sentence S from the list of novel sentences N.
   - Feature *NE-Person* counts the number of names of people recognized in each sentence S from the list of novel sentences N.
   - Feature *NE-Organization* counts the number of names of organizations recognized in each sentence S from the list of novel sentences N.
   - Feature *NE-Money* counts the number of currency expressions recognized in each sentence S from the list of novel sentences N.
   - Feature *NE-Percentage* counts the number of percentage expressions recognized in each sentence S from the list of novel sentences N.
   - Feature *NE-Date* counts the number of dates recognized in each sentence S from the list of novel sentences N.
   - Feature *NE-Time* counts the number of time expressions recognized in each sentence S from the list of novel sentences N.
   - Feature *NE-Others* counts the number of names of other entities than those quantified by the other features were recognized in each sentence S from the list of novel sentences N.
3. *Sentence length*, which was in some cases a useful predictor.

These three sets of features have the advantage that they can be generated automatically both on the training set and on the testing set. The training set was also enriched by a set of manual annotations, tagging each set with the label of one of the possible classes: H, M or L. Using the C.5 decision tree package with the boosting capability, we were able to derive multiple decision trees, used to vote the best classification. For example, when the question was *"Identify documents that discuss fatal highway accidents where the cause of the accident cannot be determined."* we obtained the following four decision trees:

```
-----  Trial 0:  -----
Decision tree:
necount <= 1: L (8)
necount > 1:
:...gloss > 0: M (3/1)
    gloss <= 0:
    :...hyper2 > 0: M (2)
        hyper2 <= 0:
        :...hyper1 <= 0: M (2)
            hyper1 > 0: H (3)

-----  Trial 1:  -----
Decision tree:
person > 0: H (3.5/1.8)
person <= 0:
:...synset <= 2: L (7.9/0.9)
    synset > 2:
    :...length <= 32: M (3.5/0.9)
        length > 32: L (3)

-----  Trial 2:  -----
Decision tree:
organization > 0: M (2.1)
organization <= 0:
:...person <= 0: L (11.8/2.9)
```

```
    person > 0: M (4.1/1.6)

-----  Trial 3:  -----
Decision tree:
necount <= 1: L (6)
necount > 1:
:...length <= 36: M (9.8/3.6)
    length > 36: L (2.1)
```

When evaluated on the training data, for the same topic we obtained the following results:

```
Trial       Decision Tree
-----      ----------------
           Size      Errors

   0          5    1( 5.6%)
   1          4    4(22.2%)
   2          3    5(27.8%)
   3          3    3(16.7%)
boost              1( 5.6%)   <<
```

We can notice that the boosted tree, obtained by voting between the four trees learned generated the same precision as the first tree, showing that for this particular answer the features provided by WordNet data are more important than the other features.

For this particular question, the decision trees selected two sentences marked as H and one sentence marked as M and produced the following 91-word summary:
"a 39-year-old Mission Viejo woman was killed Tuesday night when her car swerved into oncoming traffic on Laguna Canyon Road, police said.
John Feltman, 75, and his wife, Celia, 73, were traveling along a rural stretch of the road from their Leisure World home when the car in which they were riding drifted into oncoming traffic, Irvine Police Lt. Bob Leonard said.
Lee was driving southbound on 170th when a northbound pickup driven by John P. Hughes, 26, of San Bernandino drifted into his lane and the two collided, Sioria said.".
Figure 9 shows that the information gathered in the answers produced by our system was evaluated as being of reasonable quality, as it was ranked as the fourth system in terms of relevancy.

TASK 4 RELEVANCY		
	10	5
	13	6
	14	1
	16	3
	17	8
	19	9
	20	7
	22	**4**
	23	1

Figure 9: Relevancy evaluations for DUC-2003.

6. Evaluation

We participated with LITE-GISTEXTER in the DUC-2003 evaluations. Figure 11 lists the results of the evaluations with respect to the coverage produced by the short summaries. The Figure also shows the accuracy with which the summaries responded the twelve questions listed in Figure 10.

The results obtained by LITE-GISTEXTER were surprisingly good especially for Task 4. We interpret the results as positive in view of the very scarce semantic processing we applied and of the very few resources we employed. In contrast, the results we obtained for Task 1 show that the assumption that the first document sentence contains the most useful information is not realistic. We were also surprised by the results obtained for Tasks 2 and 3. Although the technique was the same we expected that the event-focused summaries would be of better quality than those for based on the viewpoints. The results indicate the opposite. This shows us that events need to be modeled differently in order to produce better summaries.

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 10: Questions used in the evaluations of DUC-2003.

7. Conclusions

In this paper we have shown that short summaries produced with minimal resources are of surprising quality. The simple techniques employed suggest that useful information can be captured in summaries without sophisticated semantic resources. Especially interesting is the finding that summaries that answer questions can be produced by training a classifier on very simple features, that may be annotated on new texts automatically.

8. References

M. Collins. A new statistical parser based on bigram lexical dependencies. In *Proceedings of the 34th Annual Meeting of the Association of Computational Linguistics (ACL-96)*, pages 184–191, Santa Cruz CA, 1996.

Hongyan Jing. Sentence Reduction for Automatic Text Summarization. In *Proceedings of the Conference on Applied Natural Language Processing and the North American Chapter of the Association for Computational Linguistics, (ANLP-NAACL-2000)*, 2000.

W. Kraaij, M Spitter and A. Hulth. Headline extraction based on a combination of uni- and multidocument summarization techniques. In *Proceedings of DUC-2002*, 2002.

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.

Task	Peer summarizer code baseline[1-3], manual[A-J] system [15-31])	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
TASK 1	7																				7	7		9	9
	8																				1	1		13	13
	9																				9	9		4	4
	10																				13	13		12	12
	13																				10	10		2	2
	15																				12	12		11	11
	17																				2	2		3	3
	18																				4	4		8	8
	21																				5	5		5	5
	22																				**6**	**6**		**6**	**6**
	24																				11	11		7	7
	25																				8	8		10	10
	26																				3	3		1	1
TASK 2	6	4	4	1	1	12	2	7	1	8	3	11	1	15	6						2	2		2	2
	10	6	6	8	5	1	8	5	1	12	9	11	14	13	2						12	14		10	13
	11	1	1	5	5	12	8	3	1	1	3	1	14	3	1						11	12		12	11
	12	5	3	7	5	1	2	3	11	3	12	1	8	1	11						7	5		11	6
	13	15	14	14	12	8	14	15	1	10	6	1	1	16	8						1	3		1	4
	14	13	8	13	15	12	14	13	14	3	6	1	1	3	7						6	6		3	7
	15	16	16	12	16	16	16	16	16	16	16	16	8	11	16						16	16		16	16
	16	3	5	4	1	1	2	5	1	3	1	1	1	13	9						3	4		6	3
	17	7	7	1	1	8	2	1	11	14	11	1	8	3	12						14	15		15	15
	18	12	9	9	12	12	8	7	1	15	12	1	1	6	12						9	7		8	9
	19	11	12	11	5	1	12	11	15	1	14	1	8	7	12						15	11		14	12
	20	10	15	16	5	8	1	12	1	3	2	1	1	11	5						4	1		7	1
	21	13	13	9	11	1	12	10	11	13	15	11	1	10	10						13	13		13	14
	22	**2**	**2**	**5**	**1**	**1**	**2**	**1**	**1**	**10**	**8**	**11**	**8**	**2**	**3**						**10**	**8**		**5**	**5**
	23	9	11	14	5	8	2	13	1	3	5	1	14	7	4						5	8		4	8
	26	8	10	1	12	1	8	9	1	9	9	11	8	9	15						8	10		9	10
TASK 3	10	2	2	3	6	1	2	5	1	1	4	1	5	4	2						5	5		1	4
	11	1	1	3	2	1	2	1	9	3	2	1	1	2	1						7	8		6	8
	13	9	8	9	6	6	7	10	4	7	8	1	5	8	6						10	8		10	9
	15	11	11	7	11	11	11	11	11	11	11	1	1	11	11						11	11		11	11
	16	6	7	6	6	6	10	9	4	3	5	11	1	6	6						2	6		5	7
	17	4	3	1	2	6	2	3	4	7	9	7	9	2	10						3	2		7	2
	18	5	5	5	1	1	8	3	4	9	10	7	5	1	8						4	4		4	5
	20	7	10	11	2	1	1	8	9	1	1	1	1	8	5						1	1		2	1
	21	10	6	8	6	10	2	6	4	10	5	7	11	8	8						9	3		9	3
	22	**3**	**4**	**2**	**6**	**1**	**2**	**2**	**1**	**6**	**5**	**1**	**5**	**6**	**2**						**6**	**7**		**3**	**6**
	23	8	9	10	2	6	9	6	1	3	3	7	9	4	4						8	10		8	10
TASK 4	10	1	1	2	1	1	1	1	1	4	3	1	5	5	1						8	6		6	6
	13	6	3	8	1	6	4	8	6	4	4	1	1	5	4						5	9		5	9
	14	8	7	3	5	6	8	3	1	6	5	7	9	5	5						1	1		1	1
	16	4	9	1	8	8	7	7	1	3	6	1	7	1	8						4	4		4	4
	17	6	6	6	9	1	4	5	6	7	8	1	1	3	6						6	3		8	3
	19	9	8	7	5	9	9	8	1	7	9	7	5	9	9						9	8		9	8
	20	3	4	9	1	1	2	3	6	2	1	1	1	5	6						7	6		7	7
	22	**5**	**4**	**5**	**5**	**1**	**6**	**5**	**6**	**9**	**7**	**7**	**8**	**3**	**3**						**2**	**1**		**3**	**2**
	23	2	1	4	4	1	3	2	1	1	2	1	1	2	2						3	5		2	5

Figure 11: Ranks of evaluated results in DUC-2003.

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

J. Prager, D. Radev, E. Brown, A. Coden, and V. Samn. The use of predictive annotation for question answering in TREC8. In *Proceedings of the 8th TExt Retrieval Conference (TREC-8)*, Gaithersburg, Maryland, 1999. NIST.

Rohini Srihari Cheng Liu and Wei Li. A question answering system supported by information extraction. In *Proceedings of the Conference on Applied Natural Language Processing and the North American Chapter of the Association for Computational Linguistics, (ANLP-NAACL-2000)*, pages 166–172, 2000.