LCC's GISTexter at DUC 2007: Machine Reading for Update Summarization

Andrew Hickl, Kirk Roberts and Finley Lacatusu
Language Computer Corporation
1701 North Collins Boulevard Suite 2000
Richardson, TX 75080
{andy,kirk,finley}@languagecomputer.com

Abstract

In this paper, we describe Language Computer Corporation's GISTEXTER question-focused and update-based multi-document summarization (MDS) systems. We show that by using a machine reading (MR) framework in order to construct representations of the knowledge inferable from a text collection, we were able to create coherent sets of "update" summaries that were likely to contain "new" information that could not be inferred from any previously considered document. Details of our DUC 2007 Main Task submission are provided as well.

1 Introduction

While current approaches to the task of multi-document summarization have led to the development of systems which can reliably distill the content of a collection of documents into a coherent, fixed-length summary (Blair-Goldensohn and McKeown, 2006; Conroy et al., 2006; Lacatusu et al., 2006), we believe the next generation of summarization systems – such as the "update" summarization systems evaluated as part of DUC 2007 – will need to employ dedicated mechanisms to ensure that the content of newly-generated summaries is consistent with (but not identical to) the information available from the textual resources previously available to the system. In DUC 2007, we hypothesized that a new type of natural language understanding application known as a Machine Reading (MR) system could be used in order to provide "update-based" summarization systems with robust models of the knowledge inferable from a set of documents. We assumed that by comparing the content of each update against knowledge bases (KB) compiled from sets of documents that had been previously "read" by the system, we could generate summaries that were more likely to contain "new" information that had not been previously mentioned in nor could be inferred from any previously considered document.

Following recent work in MR (Etzioni et al., 2006; Hickl and Harabagiu, 2007), we developed an MR system which leveraged state-of-the-art techniques for recognizing forms of textual inference including textual entailment (Hickl and Bensley, 2007; Hickl et al., 2006b) and textual contradiction (Harabagiu et al., 2006b) to construct representations of the knowledge encoded in a text collection. In our system, knowledge is acquired from texts by recognizing all of the available textual entailment relationships found between the set of discourse commitments extractable from the set of sentences contained in a text collection (i.e. the *text commitments*) and the set of commitments stored in the system's Knowledge Base (KB). Once a set of text commitments have been extracted, we use a state-of-the-art system for recognizing textual entailment (RTE) (Hickl and Bensley, 2007; Hickl et al., 2006b) to identify the text commitments that are entailed by at least one commitment stored in the KB. Text commitments that are not entailed by the current KB are used to update the KB prior to each new round of updates.

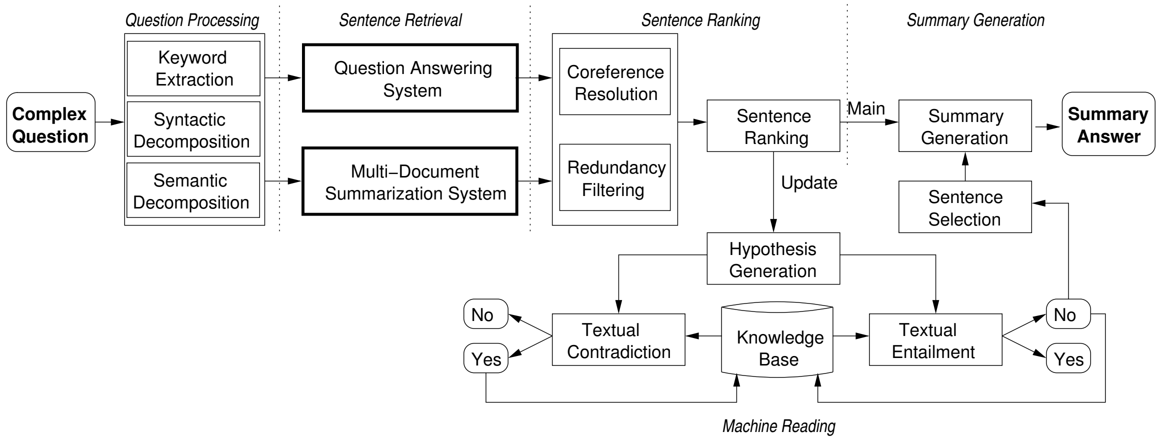

Figure 1: A Machine Reading-Based Framework for Update Summarization.

Each new update summary is generated in the following way. After a set of topic-relevant sentences has been identified, we use RTE to filter any summary sentence which textually entails at least one hypothesis in the current KB. In order to ensure that sentences that include contrasting information or corrections of previous commitments are included in update summaries, we use a system for recognizing textual contradiction (Harabagiu et al., 2006b) in order to identify any summary sentence which textually contradicts at least one hypothesis stored in the KB. Sentences which meet these two criteria are then compiled into a coherent summary using the method first described in (Lacatusu et al., 2006).

The rest of the paper is organized in the following way. Section 2 provides a detailed overview of the multi-document summarization systems developed for the DUC 2007 Main and Update tasks. Section 3 discusses results from the official DUC evaluation, and Section 4 presents our conclusions.

2 System Description

In this section, we present an overview of the systems we used to generate multi-document summaries for the DUC 2007 Main and Update Pilot Tasks. The architecture of the systems we developed is presented in Figure 1.

With GISTEXTER, question-focused multi-document summaries (such as those evaluated in the DUC 2007 Main Task) are generated in a four-step process. First, questions are submitted to a *Question Processing* module, which performs syntactic and semantic question decomposition of the complex questions created for each summarization topic. Second, subquestions identified during Question Processing are sent to a *Sentence Retrieval* module, which uses two types of sentence retrieval engines in order to identify the set of relevant sentences which should be considered for inclusion in a summary. Next, sentences are passed to a *Sentence Ranking* module, which identifies the most responsive and/or topical passages retrieved by the system. Finally, the top-ranked sentences are considered by a *Summary Generation* module, which organizes passages into coherent and responsive text passage which could be returned to a user.

The process of generating update summaries leverages the same *Question Processing*, *Sentence Retrieval*, and *Sentence Ranking* modules used by the question-focused summarization system. Once a ranked list of sentences has been assembled, the top-n most relevant sentences are then sent to a *Commitment Extraction* module, which uses a series of heuristics (similar to those used in (Hickl and Bensley, 2007)) in order to extract a subset of the discourse commitments inferable from a text passage.

These discourse commitments are then sent simultaneously to a *Textual Entailment* and *Textual Contradiction* module which evaluates whether a particular extracted commitment is consistent with – or contradicts – some commitment stored in the system's available *Knowledge Base* (KB).

Output from the system's Textual Entailment and Textual Contradiction modules are then used to both update the existing KB and to select sentences for inclusion in an update summary.

In our current model, the system's KB is assumed to represent the set of commitments that are unique to a specific set of documents. Commitments that are not textually entailed – or are textually contradicted – by the KB are automatically added to the KB after each "round" of update summarization.

Judgments from the RTE and RTC modules are sent to a *Sentence Selection* module in order to re-rank sentences based on the current state of the KB. Sentences receive a positive score for each commitment they contain that is not textually entailed by the KB. Likewise, retrieved sentences that contain commitments that are textually contradicted by a commitment stored in the KB also receive a positive score. Sentences are then re-ranked based on these judgments and are finally sent to a *Summary Generation* module for assembly into a coherent summary.

In the rest of this section, we will briefly describe each of the components used in creating both question-focused and update-based multi-document summaries.

2.1 Question Decomposition

As with our previous DUC systems (Lacatusu et al., 2005; Lacatusu et al., 2006), complex questions were initially decomposed both syntactically and semantically. Details of the syntactic decomposition techniques employed in GISTEXTER are given in (Lacatusu et al., 2005), while a full description of the algorithm used to semantically decompose questions is presented in (Harabagiu et al., 2006a). Examples of the syntactic and semantic decomposition of the complex question associated with Topic 0716D are presented in Table 1.

Original	Describe the development of Australia's uranium mine project in its Kakadu National Park and the protests and obstacles encountered.
Syntactic	Describe the development of Australia's uranium mine project in its Kakadu National Park. Describe the protests encountered. Describe the obstacles encountered.
Semantic	Who is opposed by the area's traditional owners? Where does the uranium mining lease lie? What will be removed by police? Who placed it on a list of endangered world heritage sites? Who is still doing preparatory work? What was recognized by the United Nations' World Heritage Bureau? How much do they invest in fine paper manufacturing?

Table 1: Question Decompositions for D0716.

2.2 Sentence Retrieval

As with (Lacatusu et al., 2006), we used both a question-answering system and a multi-document summarization system in order to retrieve sentences for each of the subquestions identified during question decomposition.

2.2.1 Question Answering

We used output from LCC's PALANTIR (Harabagiu et al., 2005; Hickl et al., 2006a) question-answering system in order to retrieve sentences in response to the factoid and complex questions generated during question decomposition.

Given each decomposed question, PALANTIR determined its expected answer type (EAT), which may be complex (e.g. *Describe the development of Australia's uranium mine project in its Kakadu National Park.*) or relatively simple (e.g. *Where does the uranium mining lease lie?*). For factoid questions, all entities within the document set that match the EAT were considered as answers and ranked; sentences that contained the top exact answers were considered to be relevant for inclusion in a multi-document summary. For complex questions, any sentence occurring in the top 10 ranked passages retrieved by PALANTIR for each question were considered to be relevant.

2.2.2 Multi-Document Summarization

GISTEXTER's multi-document summarization-based sentence retrieval engine uses keywords extracted from each decomposed question in conjunction with topical terms and relations computed from a set of documents in order to retrieve a relevant set of sentences that can be included in a summary.

Keywords from the decomposed questions were weighted using our question-answering keyword extraction algorithm used in LCC's PALANTIR (Harabagiu et al., 2005). Higher weights were given to proper nouns, while lower weights were given to common words. Stop words received a weight of zero.

Keywords were augmented with Topic Signature (TS_1) terms (Lin and Hovy, 2000) and Enhanced

Topic Signature (TS_2) terms (Harabagiu, 2004). (Examples of TS_1 terms are presented in Table 2.)

A hill-climbing algorithm was then used to determine the relative weight of keywords and topic signature terms. The following formula was then used to score each sentence:

$$\sum_{i=1}^{n}(h_k \log(k_i) + h_t \log(t_i))$$

where n is the number of relevant terms in the sentence, h_k is the hill-climbed weight for keywords, h_t is the hill-climbed weight for topic signature words, k_i is the keyword weight for the $i^{\mathtt{th}}$ relevant term, and t_i is the topic signature weight for the $i^{\mathtt{th}}$ relevant term.

Term	Weight	Term	Weight
uranium	491.99	world	52.12
mine	433.69	protester	48.91
australian	227.97	value	47.30
environmental	157.36	lease	44.48
park	134.09	escarpment	44.36
kakadu	106.39	floodplain	42.75
mining	82.71	square	39.96
cultural	77.80	danger	38.77
ore	76.89	area	35.73
site	69.96	report	35.08
government	66.37	waterfall	34.90
endanger	59.16	jabiluka	34.75
metric	52.25		

Table 2: Top 25 Topic Signature (TS_1) terms for Question D0716.

2.3 Sentence Ranking

Once a set of sentences had been retrieved by the Q/A and MDS systems, we used a coreference resolution system based on (Luo et al., 2004) in order to identify a candidate antecedent for every instance of a pronoun. Antecedents were inserted in the text immediately following the occurrence of a pronoun.[1])

Following coreference resolution, redundancy filtering was performed to increase the coverage of relevant information in the summary. Sentences were clustered on their key terms using k-Nearest Neighbor clustering and cosine similarity. From each cluster, the sentence containing the most information was kept for sentence scoring. (Details of the clustering method we employ in GISTEXTER are found in (Lacatusu et al., 2006).)

Each sentence retrieved by the Q/A and MDS sentence retrieval engines was then ranked based on a number of features, including (1) the relevance score derived from the sentence retrieval engine, (2) the position of the sentence in a document, (3) the length of the sentence, (4) the number of named entities or topical terms found in the sentence, and (5) the length of the document that the sentence was extracted from. A final score was determined for each sentence using a hill-climber similar to the one used in the multi-document summarization system. Sentences were then ranked according to this final score.

2.4 Summary Generation

After sentence ranking was performed, the top sentences were added to a candidate summary until the maximum length of the summary was reached. In order to ensure the creation of summaries which were locally coherent, we used a hierarchical clustering algorithm to re-order sentences that were expected to contain similar types of information. An example of one sentence cluster identified by this mechanism is presented in Figure 2.

The company wants to truck uranium ore from Jabiluka to its existing mine and plant at Ranger, about 13 miles (20 kms) away, for processing.
The company says its good environmental record at the existing Ranger mine shows the Jabiluka project poses no threat.
But the company may be left with nothing but a hole in the ground, as it goes ahead without having formal approvals to either build a processing plant or to truck ore to an established mill at its nearby Ranger mine.

Figure 2: Cluster of locally-coherent sentences (D0716).

2.5 Machine Reading

When creating update summaries, ranked sets of sentences output from GISTEXTER's Sentence Ranking module are first sent to a *Commitment Extraction* module which uses the heuristics identified in (Hickl and Bensley, 2007) in order to enumerate a subset of the discourse commitments available from each sentence[2].

[1]In order to create summaries that were shorter than the 250-word maximum length, pronouns were dropped from the summary (leaving only the resolved antecedent) when their removal would bring the total length of the summary under the word limit.

[2]Following (Gunlogson, 2001; Stalnaker, 1979), we assume that a discourse commitment (c) represents the any of the set of propositions that can necessarily be inferred to be true, given a conventional reading of a text passage.

Following (Hickl and Bensley, 2007; Hickl et al., 2006b; Harabagiu et al., 2006b), we perform the recognition of textual entailment and textual contradiction using two separate classifiers that estimate the likelihood that a text commitment is textually entailed (or textually contradicted) by one of the commitments stored in the system's available knowledge base.

Text commitments that are entailed by the KB are assumed to represent "known" information; sentences that contain entailed commitments are assigned a negative weight during *Sentence Selection*. When text commitments are not entailed by the KB, they are assumed to be "new" information that is worthy of consideration in an update summary; sentences that contain these entailed commitments are assigned a positive score at Sentence Selection, and the commitment is added to the KB. In contrast, text commitments which are textually contradicted by a KB commitment are assumed to represent "changed" information which should be included in an update summary; sentences containing these commitments are assigned a positive score during Sentence Selection, and the KB is updated to reflect the new text commitment.

3 Discussion of Results

This section presents our system's results from the DUC 2007 Main Task and Update Pilot Task.

3.1 Main Task Results

The 2007 version of GISTEXTER received slightly higher scores for all five "content"-based metrics (Content Responsiveness, Modified Pyramid, ROUGE-2, ROUGE-SU4, and BE) than the very similar system that participated in the DUC 2006 evaluation. (Results from DUC 2006 and DUC 2007 submissions are presented in Table 3.)

While our DUC 2007 system did not employ the automatic pyramid creation techniques introduced in our DUC 2006 system (Lacatusu et al., 2006), GISTEXTER's reliance on a battery of question decomposition and topic modeling techniques enabled it to produce summaries that were among the most responsive seen in this year's evaluation. While our system rank dipped slightly for the Content Responsiveness and Modified Pyramid metrics, the 2007 version of GISTEXTER received higher average scores for all five content metrics when compared to its 2006 counterpart.

Metric	**DUC 2006**		**DUC 2007**	
	Score	Rank	Score	Rank
Content Responsiveness	3.08	1	3.31	2
Modified Pyramid	0.21	4	0.31	5
ROUGE-2	0.08	12	0.11	7
ROUGE-SU4	0.14	15	0.16	7
BE	0.04	11	0.06	6
LQ1: Grammaticality	4.62	1	4.64	1
LQ2: Non-redundancy	4.60	5	3.89	6
LQ3: Referential clarity	3.71	4	4.09	1
LQ4: Focus	4.28	2	4.24	1
LQ5: Struct. and Coherence	3.28	2	3.69	1

Table 3: DUC 2007 Results for the Main Task.

ERA has government approval to build a uranium mine at its Jabiluka lease within the park's boundaries, near a uranium mine it already operates inside the park. Construction begun one week ago at Jabiluka, a uranium mining lease lying within the boundaries of Kakadu National Park. The Australian government's environmental report on the Jabiluka uranium mine (located in Kakadu Natural Park), found the area is not under threat and attacked a UNESCO report that said Kakadu Natural Park was in danger.
Many environmentalists have been calling on the Australian government not to approve the uranium mining project adjacent to the park, but the government rejected their environmental concerns, saying jobs and export income to be generated from the mining project is important. Australian conservationists and traditional aboriginal owners threatened to blockade development of the huge Jabiluka uranium mine in the country's vast Kakadu National Park. A high-level United Nations delegation on Monday began touring Australia's Kakadu National Park to examine claims a uranium mine being built in the region threatens unique cultural and environmental values. In a submission to the U.N. team, which is investigating whether the mine threatens the environmental and cultural values of Kakadu, the council said it was particularly concerned about the Malakunanja 2 archaeological site, just 1.2 miles (2 kilometers) from the mine entrance. "The mission has noted severe ascertained and potential dangers to the cultural and natural values of the Kakadu National Park, posed primarily by the proposal for uranium mining and milling at Jabiluka,".

Figure 3: Summary generated by GISTEXTER for topic D0716.

Compared to our results from DUC 2006, the 2007 version of GISTEXTER received higher linguistic quality scores for two dimensions: referential clarity (LQ3) and structure and coherence (LQ5); two other dimensions: grammaticality (LQ1) and focus (LQ4) remain essentially unchanged, while the scores for a fifth dimension, non-redundancy (LQ2) were markedly lower.

The Southern Poverty Law Center is suing **Butler**, hoping to bankrupt **him [Richard Girnt Butler]** in a lawsuit for unspecified civil damages filed in Coeur d'Alene in January.	The Southern Poverty Law Center is suing **Richard Girnt Butler**, hoping to bankrupt **him** in a lawsuit for unspecified civil damages filed in Coeur d'Alene in January.
She [Angelina Jolie] has stolen every movie **she [Angelina Jolie]** has appeared in, giving passionate performances in otherwise forgettable fare like "Hackers" and the recent "Playing by Heart."	**Angelina Jolie** has stolen every movie **she** has appeared in, giving passionate performances in otherwise forgettable fare like "Hackers" and the recent "Playing by Heart."
Arafat said **Netanyahu** does not want peace or respect the signed agreements, adding that **Netanyahu** has admitted that **he [Netanyahu]** ignores the Oslo accords.	Arafat said **Netanyahu** does not want peace or respect the signed agreements, adding that **Netanyahu** has admitted that **he** ignores the Oslo accords.
(a)	(b)

Figure 4: Example of the use of coreference information in generating summaries (a) by GISTEXTER, and (b) by a human.

Unlike our 2006 submission, we used the output of an automatic coreference resolution system in order to resolve the antecedents of all pronominal expressions found in the output summary. While this approach improved the referential clarity of our summaries, we believe that our inability to determine which pronouns should be resolved – and which should be left unresolved – was a contributing factor in our reduced non-redundancy score (DUC 2006: 4.60, DUC 2007: 3.89).

As can be seen in Figure 4, resolving all of the pronouns in a candidate summary often resulted in the creation of a text which featured the repeated mention of a name, a factor which we believe degraded the naturalness of the summary and contributed to the perception that similar kinds of information were being repeated throughout the summary.

3.2 Update Task Results

GISTEXTER obtained very competitive results across all evaluation metrics for the DUC 2007 Update Pilot Task. (Results from this year's task are presented in Table 4.[3])

GISTEXTER produced update summaries that were consistently judged to be among the top three performing systems, regardless of condition or scoring metric. The graph illustrated in Figure 5 provides a comparison of the performance of GISTEXTER with regard to the other system participating in the DUC 2007 Update Task. The graph represents the sum of the ranks obtained by the systems for each of the three types of summaries (summary A, B, or C); lower scores indicate better system performance. Figure 5 shows that GISTEXTER was markedly better on both Content Responsiveness and Modified Pyramid than the next nearest competitors (5 point difference based on Content Responsiveness rankings, and 8 point difference based on the Modified Pyramid rankings). These results suggest that forms of textual inference – such as textual inference and textual contradiction – are valuable in recognizing when information can and cannot be inferred from previous document collections. They also point out a future where summarization will depend on truly understanding the semantic content of a text, and not only rely on statistical approximations of relevance.

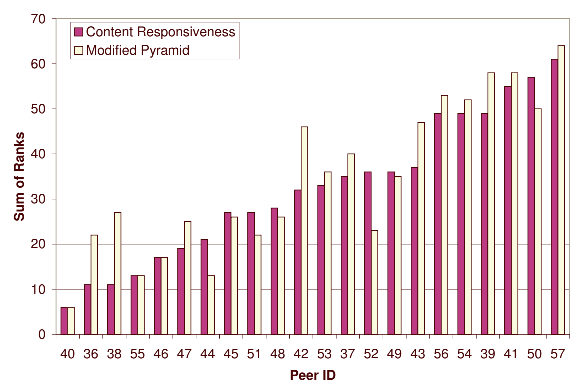

Figure 5: Sum of System Ranks for summaries A, B, and C for both Content Responsiveness and Modified Pyramid.

As with our submission to the DUC 2007 Main Task, we believe that GISTEXTER's use of a relevance model based on the output of question decomposition and topic representation systems enables it to create highly responsive summaries which are also relevant to the overall topic of a document collection. This enables us to identify the most relevant snippets of text that are also important for the

[3]The peer scores are sorted based on the average Content Responsiveness metric. R-2 and SU4 represent the ROUGE-2 and ROUGE-SU4 scores, respectively.

	Score											Rank										
	Modified Pyramid				Content Responsiveness				R-2	SU4	BE	Modified Pyramid				Content Resp.				R-2	SU4	BE
Peer	Avg	A	B	C	Avg	A	B	C	Avg	Avg	Avg	Avg	A	B	C	Avg	A	B	C	Avg	Avg	Avg
40	**0.34**	**0.40**	**0.28**	**0.34**	**2.97**	**3.30**	**2.70**	**2.90**	**0.11**	**0.14**	**0.07**	**1**	**1**	**2**	**3**	**1**	**1**	**3**	**2**	**1**	**1**	**1**
36	0.26	0.33	0.20	0.26	2.80	3.10	2.70	2.60	0.09	0.13	0.05	9	4	10	9	2	2	3	6	7	8	6
38	0.27	0.29	0.19	0.32	2.77	2.70	2.80	2.80	0.09	0.13	0.05	7	10	12	6	3	6	2	4	9	7	11
55	0.29	0.34	0.22	0.32	2.70	3.00	2.70	2.40	0.10	0.14	0.05	4	3	6	5	4	3	3	8	2	3	7
58	0.24	0.28	0.24	0.22	2.70	3.00	2.60	2.50	0.09	0.12	0.04	12	12	3	14	4	3	6	7	10	9	12
46	0.31	0.28	0.23	0.41	2.67	2.60	2.50	2.90	0.09	0.13	0.05	2	14	4	1	6	9	8	2	6	6	5
47	0.27	0.30	0.18	0.34	2.63	2.60	2.50	2.80	0.09	0.13	0.05	5	9	13	4	7	9	8	4	4	5	4
44	0.30	0.32	0.29	0.29	2.60	2.80	2.90	2.10	0.09	0.14	0.06	3	5	1	7	8	5	1	17	5	2	2
45	0.27	0.23	0.20	0.38	2.53	2.30	2.30	3.00	0.10	0.13	0.06	6	17	9	2	9	17	11	1	3	4	3
48	0.25	0.31	0.21	0.23	2.43	2.50	2.60	2.20	0.09	0.12	0.05	11	6	8	13	10	13	6	11	8	10	8
51	0.26	0.35	0.21	0.22	2.40	2.60	2.40	2.20	0.08	0.12	0.05	10	2	7	15	11	9	10	11	12	11	9
53	0.23	0.29	0.19	0.21	2.30	2.40	2.10	2.40	0.07	0.11	0.04	14	11	11	16	12	15	13	8	14	14	13
42	0.19	0.25	0.13	0.20	2.27	2.70	1.90	2.20	0.07	0.11	0.03	17	15	17	18	13	6	18	11	15	16	18
52	0.26	0.31	0.23	0.25	2.27	2.60	2.10	2.10	0.08	0.12	0.05	8	7	5	12	13	9	13	17	11	12	10
43	0.20	0.28	0.12	0.19	2.24	2.70	1.90	2.11	0.07	0.11	0.03	16	13	19	19	15	6	18	16	16	15	15
37	0.21	0.22	0.15	0.26	2.23	2.10	2.20	2.40	0.04	0.08	0.03	15	18	15	10	16	18	12	8	23	23	17
49	0.23	0.30	0.10	0.28	2.23	2.40	2.10	2.20	0.07	0.11	0.04	13	8	21	8	16	15	13	11	13	13	14
54	0.16	0.14	0.12	0.21	2.03	1.90	2.10	2.10	0.05	0.10	0.03	19	21	18	17	18	22	13	17	19	18	20
56	0.16	0.18	0.14	0.17	2.03	2.00	2.10	2.00	0.06	0.10	0.03	18	20	16	21	18	19	13	20	17	17	19
39	0.16	0.18	0.10	0.19	1.97	2.00	1.70	2.20	0.05	0.10	0.02	20	19	23	20	20	19	23	11	20	20	21
41	0.14	0.24	0.08	0.11	1.93	2.50	1.50	1.80	0.06	0.10	0.03	22	16	24	22	21	13	24	22	18	19	16
50	0.15	0.10	0.10	0.25	1.93	2.00	1.80	2.00	0.05	0.09	0.02	21	23	20	11	21	19	22	20	21	21	22
35	0.12	0.13	0.16	0.08	1.67	1.80	1.90	1.30	0.05	0.08	0.02	23	22	14	23	23	23	18	24	22	22	24
57	0.07	0.05	0.10	0.08	1.67	1.40	1.90	1.70	0.04	0.07	0.02	24	24	22	24	23	24	18	23	24	24	23

Table 4: Update Task Evaluation.

main topic of the documents. While in a 250 word summary there is space for the potential side-topics present in the document, a 100 word summary (as required by the Update Task) has space only for information relevant to the main topic. It is interesting to note that our average Content Responsiveness score for the document set A in the Update Task (3.30) was almost the same as the score we obtained for the main task (3.31). This observation sustains our previous conclusion, that GISTEXTER ranks highest the most salient information, if we take into account the fact that while the summary size decreased from 250 to 100 words, the responsiveness score remained unchanged. For document set A summary there was no need to filter out "previously reported" information, as no documents had previously been consumed by the system.

For the summaries generated from document sets B and C, the systems had to disregard information reported in the previous document sets (in set A for summaries from B, and in sets A and B for summaries generated from set C). Our responsiveness scores for these two sets, while still highly competitive, were lower than the responsiveness score for the summaries generated for document set A. The summaries generated by GISTEXTER for topic D0716D: "Jabiluka Uranium Mine" are presented in Figure 6.

4 Conclusions

This paper demonstrated how state-of-the-art systems for recognizing forms of textual inference could be used in order to generate coherent updates that maximize the amount of "new" information included in a fixed-length summary. While our DUC 2007 results were encouraging, it is important to note that the Machine Reading mechanism we have introduced here represents only one of the possible ways that textual inference information can be integrated into a multi-document summarization system. In future work, we plan to experiment with new models for knowledge acquisition which incorporate additional forms of textual inference (including temporal and model-based inference) in order to generate update summaries.

References

S. Blair-Goldensohn and K. McKeown. 2006. Integrating Rhetorical-Semantic Relation Models for Query-Focused Summarization. In *Proceedings of the Document Understanding Workshop (DUC-2006)*.

John M. Conroy, Judith D. Schlesinger, Dianne P. O'Leary, and Jade Goldstein. 2006. Back to Basics: CLASSY 2006. In *Proceedings of the Document Understanding Workshop (DUC-2006)*.

Oren Etzioni, Michele Banko, and Michael J. Cafarella. 2006. Machine reading. In *Proceedings, The Twenty-*

More than 100 protesters were arrested Friday after overwhelming police and storming onto the site of a proposed new uranium mine in Australia's north. The Australian government today gave the green light to the country's controversial uranium mining plan, arguing that it would generate billions of U.S. dollars in revenue for Australians and create 2,000 jobs. Australian conservationists and traditional aboriginal owners threatened to blockade development of the huge Jabiluka uranium mine in the country's vast Kakadu National Park, which is on the World Heritage List, after the federal government approved the mining plan for the Jabilika mine yesterday.	Archaeological sites that show humans lived in Australia up to 60,000 years ago could be damaged by a uranium mine being built within the Kakadu National Park, a high-level U.N. committee was told Wednesday. Australia faces international embarrassment over the Jabiluka uranium mine, with a U.N. committee Wednesday demanding the mine be scrapped to prevent it endangering the surrounding Kakadu World Heritage area. An old car dressed up to look like a frill-necked lizard blocked the entrance to the Jabiluka uranium mine site inside the borders of Kakadu National Park on Friday.	Many environmentalists have been calling on the Australian government not to approve the uranium mining project adjacent to the park, but the government rejected their environmental concerns, saying jobs and export income to be generated from the mining project is important. Australia's conservative government defied a U.N. body's declaration that Kakadu National Park is a World Heritage area by declaring it will allow uranium mining at Jabiluka. The Australian federal government rejected a UNESCO report which called for Kakadu in northwest Australia to be placed on the endangered list because of the threat posed by the Jabiluka uranium mine.
(a)	(b)	(c)

Figure 6: Summaries generated by GISTEXTER for the Update Task Example from document sets (a) D0716D-A, (b) D0716D-B, and (c) D0716D-C.

First National Conference on Artificial Intelligence (AAAI-2006).

C. Gunlogson. 2001. *True to Form: Rising and Falling Declaratives as Questions in English*. Ph.D. thesis, University of California Santa Cruz.

S. Harabagiu, D. Moldovan, C. Clark, M. Bowden, A. Hickl, and P. Wang. 2005. Employing Two Question Answering Systems in TREC 2005. In *Proceedings of the Fourteenth Text REtrieval Conference*.

Sanda Harabagiu, Finley Lacatusu, and Andrew Hickl. 2006a. Answering Complex Questions with Random Walk Models. In *Proceedings of the 29th Annual International ACM SIGIR*.

Sanda M. Harabagiu, Andrew Hickl, and Finley Lacatusu. 2006b. Negation, contrast and contradiction in text processing. In *Proceedings, The Twenty-First National Conference on Artificial Intelligence (AAAI-2006)*.

Sanda Harabagiu. 2004. Incremental Topic Representations. In *Proceedings of the 20th COLING Conference*, Geneva, Switzerland.

Andrew Hickl and Jeremy Bensley. 2007. Recognizing Textual Entailment with LCC's Groundhog System. In *Proceedings of the Third PASCAL Challenges Workshop (to appear)*.

Andrew Hickl and Sanda Harabagiu. 2007. Machine Reading through Textual and Knowledge Entailment. In *Proceedings of the AAAI 2007 Spring Symposium Series Workshop on Machine Reading*.

A. Hickl, J. Williams, J. Bensley, K. Roberts, Y. Shi, and B. Rink. 2006a. Question Answering with LCC's CHAUCER at TREC 2006. In *Proceedings of the Fifteenth Text REtrieval Conference (TREC 2006)*.

Andrew Hickl, John Williams, Jeremy Bensley, Kirk Roberts, Bryan Rink, and Ying Shi. 2006b. Recognizing Textual Entailment with LCC's Groundhog System. In *Proceedings of the Second PASCAL Challenges Workshop*.

F. Lacatusu, A. Hickl, P. Aarseth, and L. Taylor. 2005. Lite-GISTexter at DUC 2005. In *Proceedings of the Document Understanding Workshop (DUC-2005) Presented at the HLT/EMNLP Annual Meeting*.

Finley Lacatusu, Andrew Hickl, Kirk Roberts, Ying Shiand Jeremy Bensley, Bryan Rink, Patrick Wang, and Lara Taylor. 2006. LCC's GISTexter at DUC 2006: Multi-Strategy Multi-Document Summarization. In *Proceedings of the Document Understanding Workshop (DUC-2006)*.

Chin-Yew Lin and Eduard Hovy. 2000. The automated acquisition of topic signatures for text summarization. In *Proceedings of the 18th COLING Conference*, Saarbrücken, Germany.

Xiaoqiang Luo, Abe Ittycheriah, Hongyan Jing, Nanda Kambhatla, and Salim Roukos. 2004. A mention-synchronous coreference resolution algorithm based on the bell tree. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, pages 135–142.

Robert Stalnaker. 1979. Assertion. In Peter Cole, editor, *Syntax and Semantics*, volume 9, pages 315–332. Academic Press, London.