NeATS in DUC 2002

Chin-Yew Lin and Eduard Hovy

University of Southern California
Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292-6695
{cyl,hovy}@isi.edu
tel: 310-448-8711 and 310-448-8731

1. Overview

NeATS (Next Generation Automated Text Summarization) project focuses on multi-document summarization. The summarization strategies employed in the NeATS 2001 system including sentence position, term frequency, topic signature (Lin and Hovy 2000, 1997; Hovy and Lin 1999), term clustering, MMR (Goldstein et al. 1999), stigma word filters (Edmundson 1969), and time stamps have been proved effective in the DUC 2001 evaluation. However, the results also indicate there is still a long way to go before the current system can catch up with human performance (Lin and Hovy 2002). Analyzing the DUC 2001 results, we made the following observations:

(1) NeATS performed better in generating summaries of size of longer than 100 words but not size of shorter than 50 words.

(2) NeATS did not take advantage of domain specific information in generating summaries. For example, NeATS used the same procedure to construct biography summaries and natural disaster summaries.

(3) NeATS is an extraction-oriented system:

   a. It did not compress sentences. This is reflected in the some what worst performance in generating 50-word summaries as mentioned in (1).

   b. It did not have the capability to generalize multiple events of the same type. This often led to summaries consisting of isolated events of the same type and lacking focus.

In this paper, we describe our efforts in addressing the issues above and discuss the performance of our solutions in the context of DUC 2002. We first give a short overview of NeATS in Section 2. Section 3 describes the changes in NeATS in DUC 2002. Section 4 presents preliminary evaluation results. We conclude with lessons learned and future directions.

2. NeATS

NeATS attempts to extract relevant or interesting portions from a set of documents about some topic and to present them in coherent order. It is tailored to the genre of newspaper news articles, and it works for English, but can be made multilingual without a great deal of effort. At present NeATS produces generic (author's point of view) summaries, but it could be made sensitive to desired focus topics, input by a user.

Given an input of a collection of sets of newspaper articles, NeATS applies the following steps.

(1) Extract and rank passages — Given the input documents, form a query by building topic signatures, extract sentences, and rank them, using modules of the Webclopedia QA system (Hovy et al., 2002).

(2) Filter for content — Using an OPP policy as developed for the SUMMARIST single-document summarizer (Lin and Hovy, 1997), remove extracted sentences too far from the high-importance regions and demote sentences containing stigma words (Edmundson 1969).

(3) Enforce cohesion and coherence — Each remaining sentence is paired with a suitable introductory sentence.

(4) Filter for length — Select the required number of sentence pairs using a modified version of CMU's MMR algorithm (Goldstein et al., 1999) by comparing the similarity between existing summary as a whole instead of pair of sentences.

(5) Ensure chronological coherence — Reorder the pairs in publication order, disambiguate all time words with explicit dates, and replace relative dates with absolute dates.

(6) Format and output the final result.

3. Adjustments for DUC 2002

3.1. Use headlines for short summaries

In order to generate a 10-word summary (headline style) and pack more information into a 50-word summary, we decided to use headlines as candidate sentences for these short summaries. Headlines were not used in 100- and 200- word summaries. Only sentences in the document body were used. Since DUC 2002 required 50-word abstracts to be composed of complete sentences, we have a special routine to convert headlines in 50-word abstracts back to sentence style. Figure 1 shows an example 50-word abstract for topic 63.

```
<multi size="50" docset="d063j">
(09/22/1989) 11 Dead, 22 injured in IRA explosion.
(09/23/1989) IRA attack stirs security controversy.
(09/23/1989) 10 Killed, 22 hurt by IRA bomb at England barracks.
(09/24/1989) Prince Philip Condemns IRA, clergy urge forgiveness.
(09/25/1989) Prime minister Condemns IRA for music school explosion.
</multi>
```

Figure 1. 50-word abstract for topic 63.

3.2. Use natural disaster topic signatures and biography builder

To better process topics related to common types such as natural disasters, we built generic topic signatures (Lin and Hovy 2000). Generic topic signatures were used to supplement automatically generated topic signatures. The idea is that a generic topic signature will provide generic information for specific expected events, such as natural disasters. We manually built signatures for the following events based on DUC 2001 data: *hurricane*, *earthquake*, *drought*, *tornado*, *fire*, *volcanic*, *explosion*, and *flood*. The generic signature for *hurricane* is shown in Figure 2. If a topic was recognized as a specific type, the corresponding generic topic signature was added to the pool of dynamically generated topic signatures related to the topic. For example, if "*hurricane*" is identified as the most prominent term according to likelihood ratio λ (Dunning, 1993), then the generic *hurricane* signature is added to the pool of relevant topic signatures for the topic.

```
(:SURF "WEBCL-SUMMMARIZER-HURRICANE"
 :CAT S-NP
 :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE
 :LEX  "1"
 :SUBS
 (((HURRICANE-0)
   (:SURF "hurricane" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-LAMBDA "1"))
  ((HURRICANE-1)
   (:SURF "storm" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-3)
   (:SURF "mph" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-4)
   (:SURF "wind speed" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"
    :SUBS
    (((HURRICANE-41)
      (:SURF "wind" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMDA "1"))
     ((HURRICANE-42)
      (:SURF "speed" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMDA "1")))))
  ((HURRICANE-5)
   (:SURF "killed" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-5)
   (:SURF "killing" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-6)
   (:SURF "mile-an-hour" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-7)
   (:SURF "claims" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-8)
   (:SURF "insurers" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-9)
   (:SURF "insurance" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-10)
   (:SURF "loss" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-11)
   (:SURF "losses" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-12)
   (:SURF "rescure" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-13)
   (:SURF "caused" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-14)
   (:SURF "injuries" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-15)
   (:SURF "injuried" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-16)
   (:SURF "injuring" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-17)
   (:SURF "damage" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-18)
   (:SURF "homeless" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
((HURRICANE-19)
   (:SURF "destroyed" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
((HURRICANE-20)
   (:SURF "relief" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-21)
   (:SURF "evacuated" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-22)
   (:SURF "evacuation" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))
  ((HURRICANE-23)
   (:SURF "dead" :CAT S-NP :CLASS I-EN-WEBCL-SIGNATURE-HURRICANE :LEX "1" :-2LAMBDA "1"))))
```

Figure 2. Generic topic signature for "hurricane".

For topics of type biography, we used a biography builder, COAGENT, designed by (Hermjakob 2002) to generate a one sentence overview sentence for the topic. NeATS assumed a topic as biographical if a person name appeared more than a threshold number of times in the input texts. The threshold was determined empirically. An example biography summary is shown in Figure 3. The first line (underlined) of the summary was generated by the COAGENT. The canonical date 04/26/1989 (underlined) in the second sentence was generated by NeATS by replacing the relative date "today" from the original text.

We have tried to compress sentences to pack more information into summaries but did not have a reliable system before we submitted our final result to DUC 2002. For topics of multi-event of the same type, we did not include any special treatment this year.

```
<multi size="100" docset="d102e">
(04/26/1989) Lucille Ball, a gifted comedienne who brought laughter to millions.
(04/26/1989) Lucille Ball, the leggy showgirl, model and B-grade movie queen whose
pumpkin-colored hair and genius for comedy made her an icon of television's early
years, died early on 04/26/1989, a week after undergoing emergency heart surgery.
(04/26/1989) Miss Ball, who had a heart attack and had throat surgery in 1988,
underwent surgery at Cedars-Sinai on April 18 to replace her aorta and aortic valve
and had been getting out of bed, eating and even walking around the room in recent
days.
(04/27/1989) A private burial was planned, reportedly with no funeral services in
accordance with Miss Ball's wishes.
</multi>
```

Figure 3. Example biography summary for topic 102, "Lucille Ball".

4. Results

The preliminary results shown in Figure 4 indicate that NeATS (system 26) does fine in the DUC 2002 evaluation. It ranks 2nd in summaries of 10-, 50-, and 200-word in coverage but does badly in summary of 100 words. We are investigating the reasons for this abnormality. NeATS also performs well in providing relevant information in its summaries that was not presented in human summaries. Almost half of the content of summaries produced by NeATS are relevant (53.2% for 50-word, 57.6% for 100-word, and 65.8% for 200-word). However, NeATS seems to suffer in quality; whether this is due to the use of headlines or replacement of relative dates with absolute dates is subjected to further study.

According to the results, the human (XXX-H) is still better than the best system (system 19) by 25 to 30 percent. One of the baseline systems (XXX-3) performs well in long summaries (100- and 200-word). System rankings, according to coverage and length adjusted coverage scores, do not differ much. This might be due to DUC 2002 participants honoring the length restrictions and at the same time also trying to provide as much information as they were allowed to. Overall, we believe DUC 2002 is a worthy exercise and would like to see evaluations such as this to continue in the coming years.

5. References

Dunning, T. 1993. Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics* 19, 61–74.

Goldstein, J., M. Kantrowitz, V. Mittal, and J. Carbonell. 1999. Summarizing Text Documents: Sentence Selection and Evaluation Metrics. *Proceedings of the 22nd International ACM Conference on Research and Development in Information Retrieval (SIGIR-99)*, Berkeley, CA, 121–128.

Edmundson, H.P. 1969. New Methods in Automatic Abstracting. *Journal of the Association for Computing Machinery*. 16(2).

Hermjakob, U. 2002. COAGENT: Constructing Complex Answers to Short Questions Using Natural Language Relevance Patterns. In preparation.

Hovy, E.H. and C.-Y. Lin. 1999. Automated Text Summarization in SUMMARIST. In M. Maybury and I. Mani (eds), *Advances in Automatic Text Summarization*. Cambridge, MA: MIT Press.

E.H. Hovy, U. Hermjakob, Chin-Yew Lin, and Deepak Ravichandran. 2002. Using Knowledge to Facilitate Pinpointing of Factoid Answers. In *Proceedings of the 19th International Conference on Computational Linguistics* (COLING 2002), Taipei, Taiwan, August 24 - September 1, 2002.

Lin, C.-Y. and E.H. Hovy. 1997. Identifying Topics by Position. *Proceedings of the 5th Conference on Applied Natural Language Processing* (ANLP). Washington, DC.

Lin, C.-Y. and E.H. Hovy. 2000. The Automated Acquisition of Topic Signatures for Text Summarization. *Proceedings of the COLING Conference*. Saarbrücken, Germany. August, 2000.

Lin, C.-Y. and E.H. Hovy. 2002. From Single to Multi-document Summarization: A Prototype System and its Evaluation. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics* (ACL-02), Philadelphia, PA, U.S.A., July 7-12, 2002.

Size-System	Instances	Rank	Avg-Coverage	Adj-Rank	Adj Avg-Coverage	Relevant	Quality
10-Word							
010-16	59	6	0.120	6	0.080	0.153	0.000
010-19	59	2	0.390	2	0.260	0.000	0.000
010-20	59	5	0.127	5	0.085	0.119	0.000
010-25	59	4	0.137	4	0.091	0.136	0.000
010-26*	59	3	0.255	3	0.170	0.102	0.000
010-29	59	7	0.091	7	0.061	0.119	0.000
010-H	58	1	0.494	1	0.329	0.052	0.000
50-Word							
050-16	59	9	0.123	10	0.082	0.231	0.647
050-19	59	2	0.234	2	0.159	0.376	0.461
050-2	59	7	0.138	7	0.097	0.400	0.420
050-20	59	10	0.122	8	0.096	0.298	0.783
050-24	59	8	0.129	6	0.114	0.237	0.593
050-25	59	11	0.100	11	0.067	0.512	1.295
050-26*	59	3	0.207	3	0.139	0.532	1.018
050-28	59	4	0.206	4	0.138	0.136	0.684
050-29	59	5	0.161	5	0.123	0.410	0.754
050-3	59	6	0.140	8	0.096	0.224	0.590
050-H	58	1	0.361	1	0.241	0.566	0.397
100-Word							
100-16	59	9	0.128	10	0.085	0.315	0.981
100-19	59	2	0.235	2	0.158	0.532	0.736
100-2	59	10	0.126	11	0.084	0.529	0.565
100-20	59	7	0.160	7	0.108	0.498	1.004
100-24	59	5	0.189	5	0.131	0.410	0.897
100-25	59	11	0.122	9	0.087	0.556	1.260
100-26*	59	6	0.186	6	0.124	0.576	0.909
100-28	59	3	0.228	3	0.152	0.332	0.879
100-29	59	8	0.142	8	0.096	0.536	1.008
100-3	59	4	0.204	4	0.136	0.458	0.894
100-H	58	1	0.341	1	0.228	0.607	0.388
200-Word							
200-16	59	10	0.151	10	0.101	0.475	0.982
200-19	59	2	0.254	2	0.170	0.634	0.898
200-2	59	11	0.128	11	0.086	0.749	0.603
200-20	59	7	0.206	7	0.138	0.668	1.033
200-24	59	6	0.218	6	0.147	0.546	0.989
200-25	59	9	0.156	9	0.105	0.603	1.243
200-26*	59	3	0.243	3	0.162	0.658	1.050
200-28	59	4	0.228	4	0.152	0.468	1.058
200-29	59	8	0.166	8	0.112	0.661	1.086
200-3	59	5	0.220	5	0.154	0.495	0.994
200-H	58	1	0.364	1	0.243	0.800	0.576

Figure 4. Summary of DUC 2002 evaluation on multi-document abstracts (Avg-Coverage: average coverage, Adj-Rank: length adjusted rank, Adj Avg-Coverage: length adjusted average coverage, Relevant: percentage of unmarked peer units that are relevant, Quality: average of how many non-zero quality questions were marked per topic).