Presenting Extracted Information in Multidocument Hypertext Summaries
(Demo Submission for HLT ’02)

Michael White,1 Claire Cardie,2 Vincent Ng2 and Daryl McCullough1

1CoGenTex, Inc.
840 Hanshaw Road
Ithaca, NY 14850, USA
mike,daryl@cogentex.com

2Department of Computer Science
Cornell University
Ithaca, NY 14850, USA
cardie,yung@cs.cornell.edu

Demo Description

The RIPTIDES system combines information extraction and multidocument summarization to produce domain-specific hypertext summaries of a set of news articles on the same event. The hypertext summaries include a high-level textual overview, tables of extracted facts, and targeted access to supporting information from the original articles. The overview uses sentences extracted from the input articles to convey the most important information about the event. Figure 1 shows a sample textual overview for a powerful earthquake that took place in Afghanistan in 1998, using the articles from TDT topic 89 as input. The indexed set of tables present the extracted facts in semantically comparable groups. Thus, for example, estimates of the overall death toll are separated out from estimates of the death toll in specific locales, or the death toll due to previous quakes in the same area. Figure 2 shows a sample page of tables displaying all the extracted estimates of the overall death toll.

In this demo, we will show how the hypertext summaries can aid an information analyst in identifying discrepancies in numeric estimates across news sources. In a companion poster submission, we describe a case study whose results suggest that, on average, the hypertext summaries provide a significantly more complete picture of the available information than the latest article. While the analyst must examine the hypertext tables carefully in order to separate out the best available estimates from those that have been superceded or are incorrectly classified, in our experience we have found that it is much easier to scan the tables for this information than to scan the full text of the original articles.
Relief workers in earthquake-stricken Afghanistan say their biggest challenge is getting food to hungry survivors. The workers say many in the northern provinces of Takhar and Badakhshan have not eaten since the quake hit on Saturday.

Workers are having a difficult time reaching victims because landslides have cut off roads to remote villages. Up to 10,000 people were killed; thousands more are missing or homeless.

Aid workers have airlifted dead bodies of victims to Idas in Afghanistan. Food, water, medicine and other supplies have started to arrive.

The United Nations, the Red Cross, and several agencies have turned to helicopters to deliver aid.

**Figure 1: Sample Textual Overview**

<table>
<thead>
<tr>
<th>Date</th>
<th>Source</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>06/02/1998</td>
<td>CNN</td>
<td>The European commission says it's sending $1,600,000 -- almost $2 million -- in relief to victims of an earthquake in Afghanistan. Food, water, medicine and other supplies have started to arrive. But a U.N. relief coordinator says it's a &quot;scenario from hell.&quot; Workers are having a difficult time reaching victims because landslides have cut off roads to remote villages. Up to 4,000 people were killed. Thousands more are missing or homeless.</td>
</tr>
</tbody>
</table>

**Figure 2: Sample Tables of Comparable Extracted Facts**