

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

Michael White,<sup>1</sup> Claire Cardie,<sup>2</sup>  
Vincent Ng<sup>2</sup> and Daryl McCullough<sup>1</sup>

<sup>1</sup>CoGenTex, Inc.  
840 Hanshaw Road  
Ithaca, NY 14850, USA  
`mike,daryl@cogentex.com`

<sup>2</sup>Department 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examine the hypertext tables carefully in order to separate out the best available estimates from those that have been superseded or are incorrectly classified, in our experience we have found that it is much easier to scan the tables for this information than to scan the full text of the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.

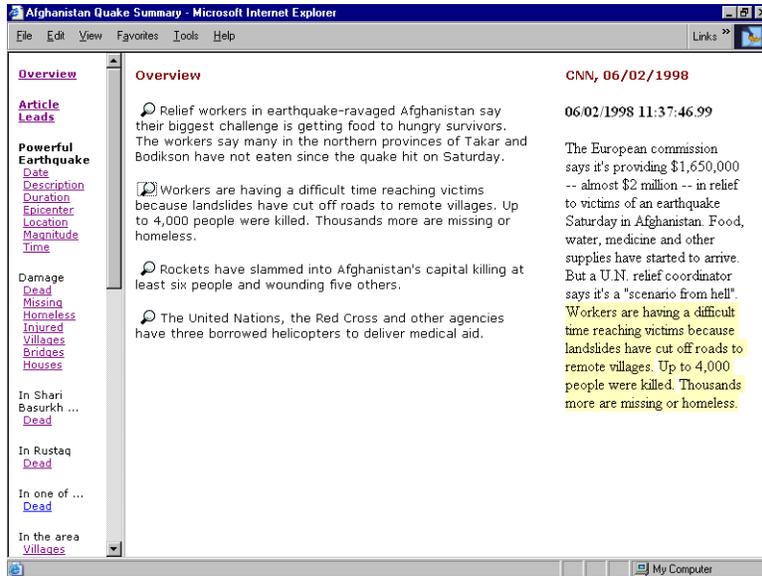


Figure 1: Sample Textual Overview

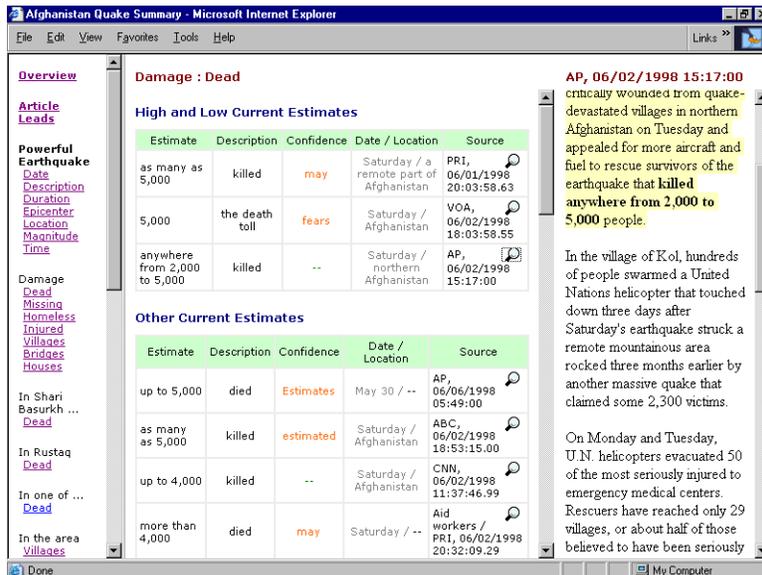


Figure 2: Sample Tables of Comparable Extracted Facts