TAP-XL: An Automated Analyst's Assistant

Sean Colbath, Francis Kubala

BBN Technologies, 10 Moulton Street, Cambridge, MA 02138

scolbath@bbn.com, fkubala@bbn.com

Abstract

The TAP-XL Automated Analyst's Assistant is an application designed to help an English-speaking analyst write a topical report, culling information from a large inflow of multi-lingual, multimedia data. It gives users the ability to spend their time finding more data relevant to their task, and gives them translingual reach into other languages by leveraging human language technology.

1 System Description

The TAP-XL system exploits language technology to monitor the user's interactions with the system and provide suggestions of relevant information to analysts, maximizing time spent reading relevant documents and writing reports. Any document, passage, or fact that a user saves in a report is deemed to be valuable, and the TAP-XL system then proactively suggests related information that is located in the stream of documents. Rather than force the user to learn a suite of distinct tools, this "suggestion" metaphor is employed throughout TAP-XL: all language technologies pull together to bring value to the user through data, rather than via additional tools, interfaces, and metaphors that must be learned separately.

2 Use Model

The user interacts with the TAP-XL system via a traditional word processor (Microsoft Word), with an additional web-based user interface.

The user writes a report based on an initial problem statement. The problem statement includes names of people, locations, and organizations relevant to a topic, as well as hypotheses about events involving these entities that are to be corroborated or refuted in the report.

Suggestions appear in a window to the left of the word processor. The initial set of suggestions is generated from the text of the problem statement. Each suggestion leads to a document or collection of documents. Passages from a document deemed relevant by the user can be cited in the report via a "create citation" button. This places the selected excerpt in the report, along with a hyperlink to the original source document. It also triggers the TAP-XL system to provide additional suggestions relevant to the passage the user selected, as well as documents relevant to any entities included in the citation. Suggestions deemed not relevant can be deleted. A screenshot of the TAP-XL system in use may be seen in Figure 1.

In addition to the suggestion mechanism, users may employ a traditional keyword-based query mechanism to locate documents in the system.

The process of system suggestion $\rightarrow$ citation $\rightarrow$ additional suggestions results in a feedback loop between the TAP-XL system and the user's report. This feedback loop is designed to allow many different human-language technologies to contribute relevant information to the user.

3 System Architecture

The analytical portion of TAP-XL is a distributed, component-based system. Components include Machine Translation (Arabic to English), Document Clustering, Multi-document Summarization, and Fact Extraction.

The components are distributed across the Internet, using a custom web service technology called the TAP Connector. The TAP Connector uses industry-standard web protocols to communicate between a requester and a provider, allowing distributed computation across the Internet with unpredictable data flows and latencies.

All metadata produced by the components is stored in a central data repository, making it available to other downstream technologies.

The system currently processes approximately 1,000 English newswire documents per day from a commercial source, as well as 150 documents per day from Arabic newspapers, obtained via web harvesting. All Arabic documents are translated to English via Machine Translation. The total flow of documents is then exposed to all the other downstream technologies (Fact Extraction, Document Clustering, and Multi-document Summarization).
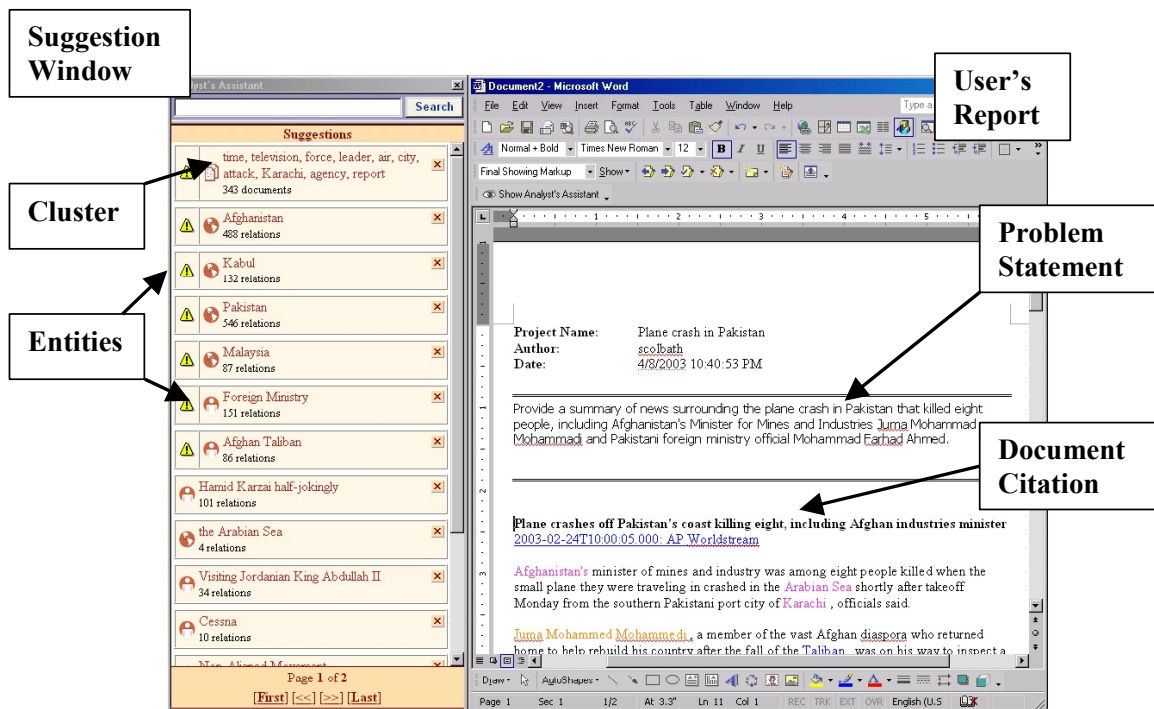
Figure 1. TAP-XL System in Use

4 Recent Developments

The TAP-XL system has been under development since November of 2002, and was recently used by more than 12 analysts in an Integrated Feasibility Experiment (IFE) under the DARPA TIDES program. The results of this IFE will be used to guide future enhancements of the TAP-XL system, including user interface enhancements, use model improvements, and additional language component technology.

5 References

Francis Kubala, Sean Colbath, Daben Liu, Amit Srivastava, and John Makhoul. 2000. Integrated technologies for indexing spoken language. *Communications of the ACM*. February, 43: 48-56.

John Makhoul, Francis Kubala, Timothy Leek, Daben Liu, Long Nguyen, Richard Schwartz, and Amit Srivastava, 2000. Speech and language technologies for audio indexing. *Proceedings of the IEEE*, 88:1338-1353.

Simon, Herbert A. 1995. *Knowledge and the Time to Attend to It*. Working paper no. 96-2, Carnegie Bosch Institute for Applied Studies in International Management, Carnegie Mellon University, Graduate School of Industrial Administration.