

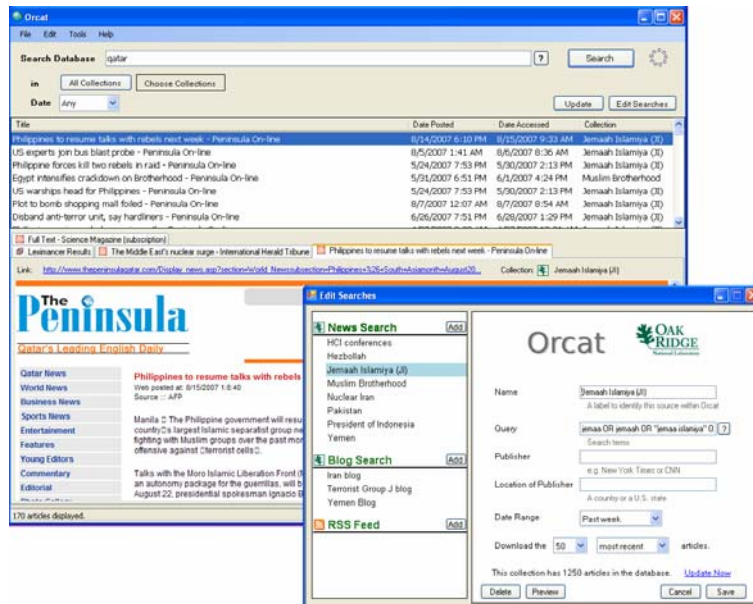


# ORCAT

## ORNL Content Analysis Tool



According to a 2003 study of the Joint Forces Intelligence Command, seventy percent of an analyst's time is spent in data management tasks. While the value of open source intelligence is increasingly recognized, its heterogeneous nature, volume, and inefficient access only amplify the problem.



**Approach:** ORCAT's approach is to assist the analyst by automating repetitive data management tasks. ORCAT feeds *open source Web content*, such as news items, reports, bulletins, multi-media, and blogs, into an automated pipeline. Items' source and date are preserved. ORCAT's functionality includes storage, display, search and analysis. It relies on RSS (Real Simple Syndication), a Web 2.0 technology and Leximancer, a COTS proven Content Analysis algorithm to assist in the performance of daily intelligence tasks. Using ORCAT, the analyst may:

- Automatically build and store customized collections of news media and blog content in real time.
- Seamlessly search the Web and personal collections using two-tier searches for improved precision.
- Display items of interest in the context of their source and on the desktop.
- Quickly provide a map of the main topics in their context.
- Follow the evolution of trends with regard to a particular subject matter.
- Profile a topic of interest in a given collection or across collections.

**Uniqueness:** ORCAT presents a single, smooth, and intuitive user interface while providing the functionality of many tools: automatic storage of Web search results, collection building, content harvesting, multi-tier and set searches, data clean-up, metadata, unsupervised content analysis, and data visualization. It is a lightweight, desktop tool allowing the analyst to exercise the power of human judgment at crucial steps. Customized collections are built from the Web and/or preferred sources without the costs of proprietary content. The Leximancer algorithm produces a ranked summary of essential concepts in a collection within the context of a few sentences. As ORCAT is built with components in an open architecture, other tools may be used within its framework.

**Points of Contact:** Line C. Pouchard, PhD. Joe Trien., Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, TN 37831-6367. Email: [pouchardlc@ornl.gov](mailto:pouchardlc@ornl.gov), [trienjp@ornl.gov](mailto:trienjp@ornl.gov), telephone (865-574-6125).