

IBM Introduces Open Software Framework for Analysis of Unstructured Information

by Paula J. Hane

August 15, 2005 — IBM has been working on an engineering foundation to support applications that make sense of the facts, meaning, and relationships in unstructured information. IBM's Unstructured Information Management Architecture (UIMA) is an architecture and framework platform for creating, integrating, and deploying unstructured information management solutions that use combinations of semantic analysis and search components. The company has just announced it is making UIMA available as free, open source software. The UIMA framework has already been embedded in IBM products, including IBM WebSphere Information Integrator OmniFind Edition, which IBM said is the first commercially available software platform for processing content based on the UIMA standard. IBM WebSphere Portal Server and Lotus Work Place also leverage UIMA for content processing. In addition, 16 software vendors have announced commercial support for and adoption of UIMA.

According to information on the IBM Research site, an Unstructured Information Management solution generally may be characterized as a software system that analyzes large volumes of unstructured information (text, audio, video, images, etc.) to discover, organize, and deliver relevant knowledge to the client or application end user. An example is an application that processes millions of medical abstracts to discover critical drug interactions. Another example is an application that processes tens of millions of documents to discover key evidence indicating probable competitive threats.

"UIMA provides, for the first time, true interoperability among different knowledge discovery, search, business intelligence, and text analytics software," noted Arthur Ciccolo, department group manager for information and knowledge management, IBM Research. "This initiative will enable organizations to deliver groundbreaking solutions that can leverage unstructured information in entirely new and advanced ways."

Hadley Reynolds, senior analyst and director of research for the Delphi Group, clarified in a NewsFlash exactly what IBM's UIMA framework provides: "[It] proposes a new 'standard' for text analytics implementations that includes common interface definitions and a common data model. It does not include a search engine for distribution or a runtime environment in which to process and provision analytic applications to business systems." To do these functions, he said, customers can license the IBM OmniFind Edition. Hadley said the big news is that IBM is throwing its weight behind an infrastructure that can reduce the complexity of implementing analytics applications.

Consultant Dana Gardner, a former analyst with the Yankee Group, called the announcement "the latest example of IBM's apt balancing of openness and opportunity." He also noted the benefits of a cooperative approach: "UIMA-compliant analytics tools, which companies will still have to purchase and support, by the way, can gang-tackle the search problem, rather than expect one tool to do it all well." He wrote optimistically about the possibilities: "Should it quickly gain ground as appears the initial case, UIMA can significantly help close the gap between tacit human knowledge and what search engines do so well, namely to index and match labels. UIMA can take enterprise search to the next level. ..."

UIMA is the result of more than 4 years of development by IBM Research. It received significant support from the Defense Advanced Research Projects Agency (DARPA) and from several universities, along with industrial research and development organizations. Some of the universities that participated, such as Carnegie Mellon University, Columbia University, Stanford University, and the University of Massachusetts Amherst, are already using UIMA in courses and research projects. The other organizations actively supporting and using UIMA include Science Applications International Corp., BBN Technologies, The Mayo Clinic, and MITRE Corporation.

Analyst John Blossom of Shore Communications, Inc. commented: "It's great news for companies specializing in content analytics, offering them an easy integration framework that can accelerate the ROI cycle for their offerings, but perhaps somewhat less rosy news for Microsoft and other integrators who are just beginning to pick up the scent on mining unstructured content's full value."

One of the companies that has announced its support of UIMA is Factiva. According to Greg Gerdy, vice president and director of channel marketing and strategy, Factiva is currently working to develop a "data listener," a sort of binding layer that will allow Factiva content to be used by UIMA-compliant applications. In addition, Factiva is working with Attensity and ClearForest, two UIMA-compliant software companies, to develop role-specific applications for government agencies and for sales, respectively. The companies expect to have proof-of-concept applications to show to potential customers later this year.

Gerdy said there's currently a lot of activity in the text analytics space. He feels the market will benefit by having a company of the stature of IBM driving developments. He commented: "IBM's framework equips Factiva and its UIMA-compliant partners to serve key roles and multiple vertical markets."

Links

The UIMA Software Development Kit can now be downloaded free of charge from IBM AlphaWorks at <http://www.alphaworks.ibm.com/tech/uima>.

The UIMA technology will be presented to the Open Source Technology Group (<http://www.ostg.com>). Availability through SourceForge (<http://sourceforge.net>) is expected by the end of 2005.

More information about WebSphere Information Integrator OmniFind Edition can be found at http://www.ibm.com/software/data/integration/db2ii/editions_womnifind.html.

To read "Dana Gardner parses IBM's corporate search standard," go to <http://blogs.zdnet.com/BTL/?p=1690>.

To access the Delphi Group NewsFlash "IBM Opens Push for Text Analytics Architecture Standard," go to http://www.delphiweb.com/knowledgebase/newsflash_guest.htm?nid=982.

Companies with Plans to Support UIMA

Attensity: <http://www.attensity.com>

ClearForest: <http://www.clearforest.com>

Cognos: <http://www.cognos.com>

Endeca: <http://www.endeca.com>

Factiva: <http://www.factiva.com>

Kana: <http://www.kana.com>

InQuira: <http://www.inquiria.com>

iPhrase: <http://www.iphrase.com>

Inxight: <http://www.inxight.com>

Nstein: <http://www.nstein.com>

QL2: <http://www.ql2.com>

SAS: <http://www.sas.com>

SchemaLogic: <http://www.schemalogic.com>

Semagix: <http://www.semagix.com>

SPSS: <http://www.spss.com>

TEMI S: <http://www.temis.com>

Paula J. Hane is Information Today, Inc.'s news bureau chief and editor of NewsBreaks. Her e-mail address is phane@infotoday.com.

[Make infotoday.com your home page](http://www.infotoday.com)