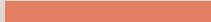


Introduction



The opportunity presented by unstructured information management (UIM) is enormous, and we believe it will significantly change not only our industry but society as well over the next decade. Database technology fundamentally changed the way our customers did business, enabling the management of every aspect of their operations in a coordinated and integrated fashion through access to structured data. UIM will extend that impact by unlocking the value that is captured but underutilized in unstructured information.

Business documents within the enterprise, syndicated sources of industry information, and the wealth of information on the World Wide Web all share the characteristic of being predominantly unstructured; that is, the content is mainly text, image, video, and audio. Emerging technologies that allow computers to directly process content and extract meaning from it will enable new types of solutions to societal and business problems.

Imagine a pharmaceutical firm that can automatically mine the text in millions of technical journal articles, patent applications, and clinical records to find evidence that a particular disease may be inhibited by a proposed treatment. Such a capability could dramatically lower the cost and reduce the time to bring new drugs to market. Or consider a manufacturing firm that automatically mines the text of millions of lines of correspondence, e-mail, and call-center transcribed notes to find incipient problems with products so that it can quickly intervene before these problems mushroom out of control. Such a firm would protect both its customers and its reputation.

These capabilities are now emerging from the research labs and being employed by our leading edge customers to address a multitude of business opportunities. As the technologies and the means for delivering them mature, new solutions that merge the value in structured and unstructured information processing will become ubiquitous.

In this issue, examples of the core technologies used to extract information from unstructured sources are presented, as well as a number of papers focused upon the means for more rapidly delivering these technologies into customers' applications. A major theme here is the utilization of IBM's UIMA (Unstructured Information Management Architecture) to integrate the capabilities of state-of-the-art search engine technology with advanced text analytics. It is our belief that UIMA will play a fundamental role in helping IBM and its business partners and customers to author and integrate new analytic capabilities within a scalable, componentized

infrastructure. It will also stimulate the research community towards greater advances in unstructured information processing—advances that will arise from a growing ability to integrate a collection of diverse analytic-processing techniques and to use the community's collective capabilities to provide results of much higher quality.

With this issue, we welcome you to the next decade of information-integration technology.

A handwritten signature in black ink, appearing to read "Janet Perna". The script is fluid and cursive, with the first name being more prominent.

Janet Perna
General Manager
Data Management
Software Group

A handwritten signature in black ink, appearing to read "Alfred Spector". The signature is highly stylized and cursive, with a large, sweeping initial letter.

Alfred Spector
Vice President
Software & Services
Research Division