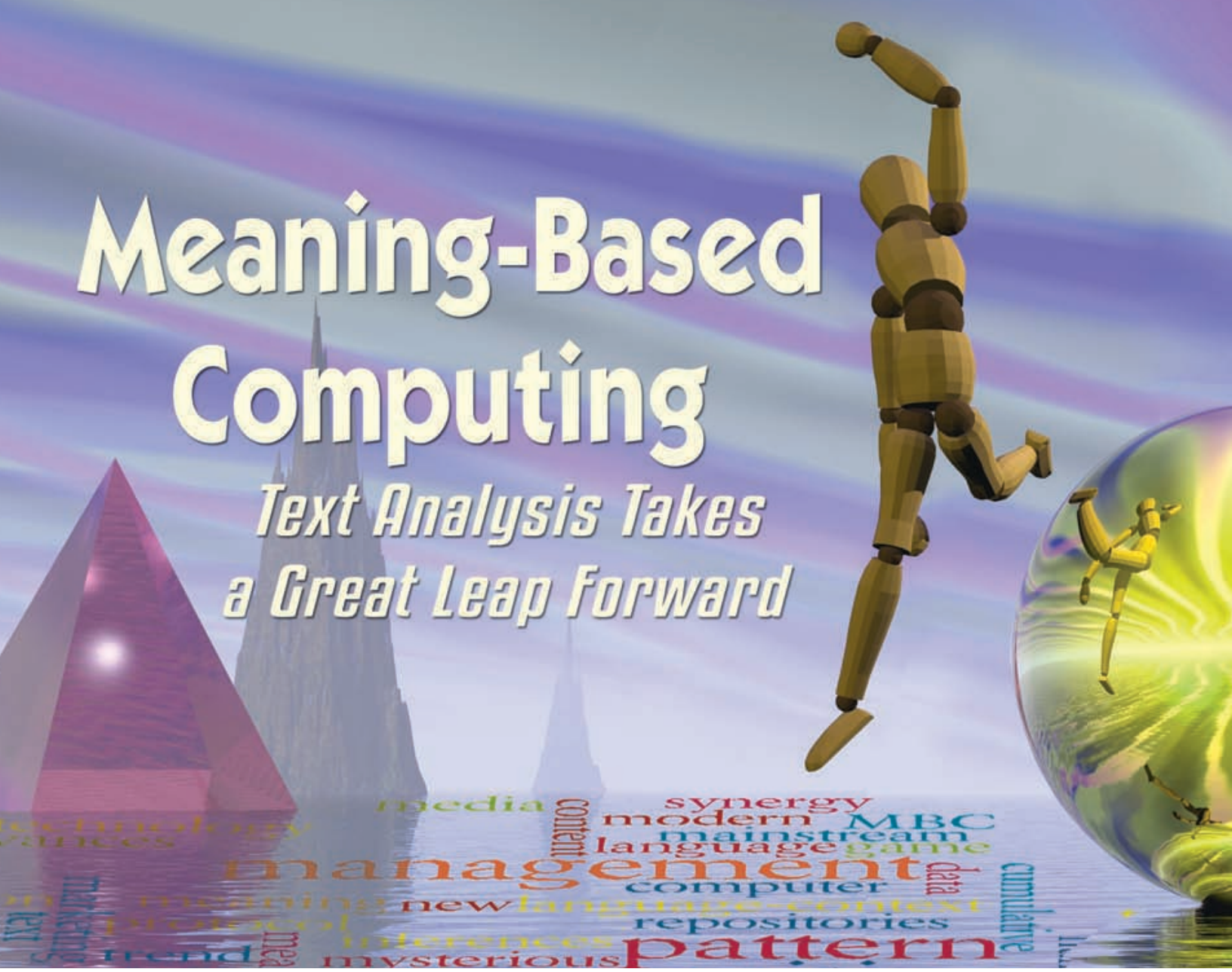


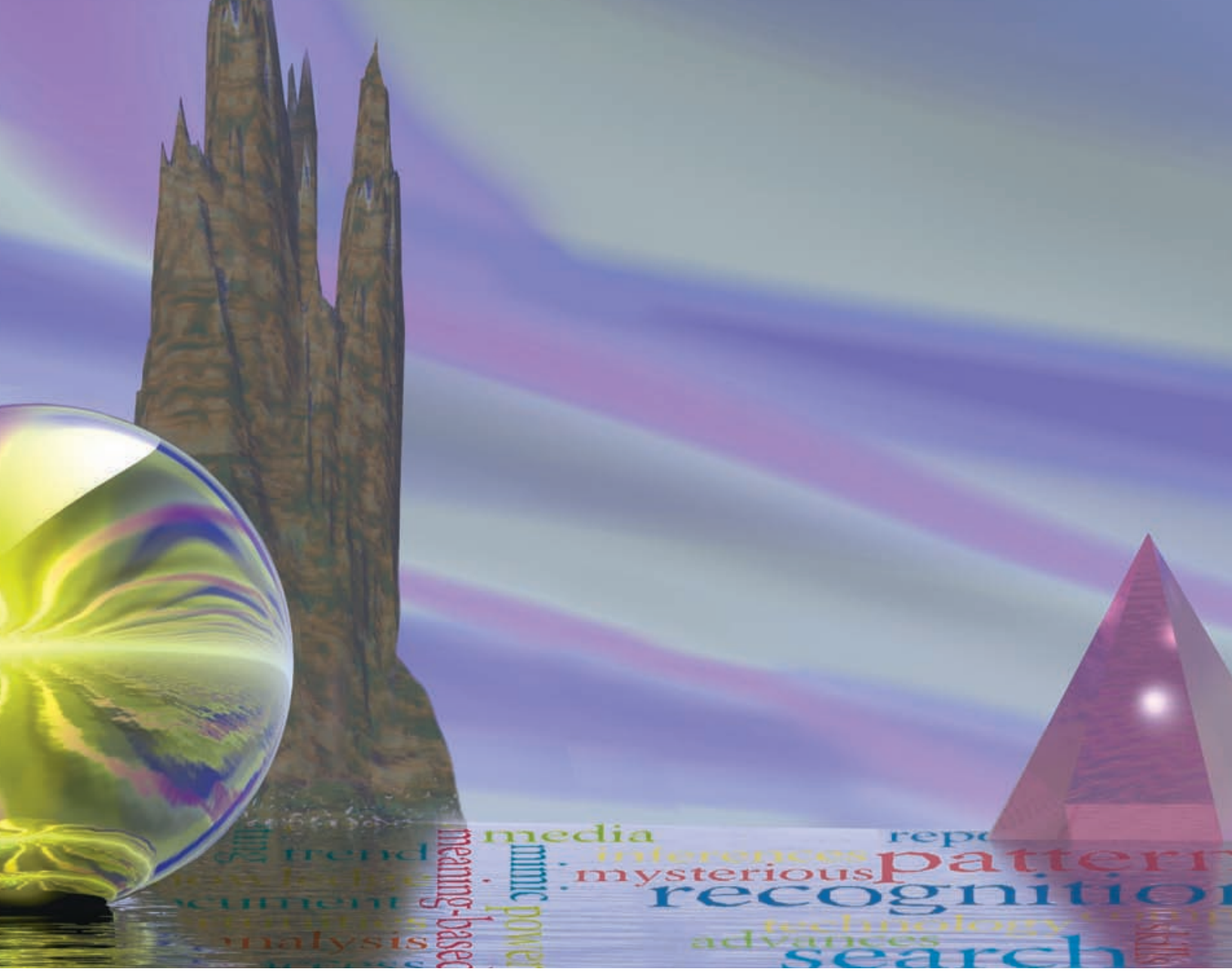
HP announced on Aug. 18, 2011, that it was buying Autonomy. —Ed.

Calling new technology game changing sounds like marketing hype. But every now and then, it's actually true. In the knowledge management area, I would cite meaning-based computing (MBC) as the new(er) technology that will shift how businesses operate. The formative events of MBC have occurred under the radar of the mainstream media, although they caught the attention of the computer press as far back as 2006, when IDC named MBC one of the top trends in content management, search, and access technologies in a report of the same name written by Susan Feldman, Melissa Webster, Abner Germanow, and Joel N. Martin.

An article by John Markoff, "Armies of Expensive Lawyers, Replaced by Cheaper Software," published in *The New York Times* on March 4, 2011, online and in the March 5, 2011, printed newspaper (www.nytimes.com/2011/03/05/science/05legal.html), surfaced the capabilities of this enterprise-level software to a more general readership.

MBC unites the power of modern search protocols with recent advances in text pattern recognition, language-context analysis, and even "sentiment analysis"—which sounds somewhat mysterious. The cumulative advances in MBC are enabling computers to make far more useful inferences about the meaning of communications, even as language usage evolves. Although MBC has been used over the past 8 years with considerable success, the synergy of recent advances has gained wider attention.

While MBC demonstrates new value for enterprise computing, should information professionals be interested in it? The short answer is yes. Whenever computers learn to mimic human skills in pattern recognition and to make inferences about the meaning of language, every reference provider should take note. The longer answer is more nuanced and intriguing, because disruptive technologies usually have an outsized impact on professional work. If MBC can in fact discover enhanced meaning in data sets



by Terence K. Huwe

and document repositories, it will also generate new opportunities for innovation. With that in mind, a review of MBC and its recent breakthroughs follows, along with its historical roots. Three strategies are offered in conclusion to challenge readers to get ready for the new frontier that MBC opens up.

THE 'CONTEXT' OF CONTENT

MBC is the brainchild of University of Cambridge's Michael Lynch, who founded Autonomy, a global consultancy that uses MBC to increase productivity, to mitigate the risk of lost data, and to improve strategic planning. In studying search technologies, Lynch came to realize that data warehousing could provide access to vast repositories of both *structured* and *unstructured* data, a very important distinction. He concluded that prevailing search techniques were not keeping pace with the migration of data from structured formats (such as databases) to unstructured formats (email, telephone conference calls, documents on disorganized directory trees, and so on). He estimated that unstructured data now account for as much as 85% of the total data we use—and that much of this content cannot be recovered by standard techniques.

Lynch saw a need for search protocols that could extract meaning from data in both structured and unstructured formats, uncovering linkages between diverse documents, associating and analyzing the meaning of words in various contexts, and ultimately discovering the intentions of the writers. "Big data" made serious experimentation possible, as business firms and government agencies started managing vast amounts of information. These mega-troves presented computer scientists with ideal test beds for discovering meaning through automated analysis.

Although MBC is a new application, the theory that inspired it dates from the mid-1740s. Lynch drew inspiration from Bayes' theorem, a mathematical concept that

Resources

Markoff, John. "Armies of Expensive Lawyers, Replaced by Cheaper Software." *The New York Times*, March 5, 2001, p. A1

Autonomy
www.autonomy.com

Thomas Bayes Wikipedia entry
http://en.wikipedia.org/wiki/Thomas_Bayes

IT Toolbox. Meaning-Based Computing
http://it.toolbox.com/wiki/index.php/Meaning-Based_Computing

Pollock, Ellen. "The Eight Million Dollar Associate." *The American Lawyer*, Vol. 13, No. 4, May 1986

Silberman, Steven. "The Quest for Meaning." Wired 8.02
www.wired.com/wired/archive/8.02/autonomy_pr.html

Singh, Rajiv. "Meaning-Based Computing: A Broad Church." Domain-b.com, Sept. 4, 2006
www.domain-b.com/infotech/itfeature/20060904_computing.htm

explains the probability of things happening, including the concept of "inverse probability." Thomas Bayes, a British cleric and scholar, was seeking a mathematical basis for proving the existence of God, but Lynch saw an entirely different potential for Bayesian theory. When applied to information retrieval, Bayesian modeling essentially places a rocket booster under pattern recognition protocols, enabling computers to make qualified inferences about related concepts and terms that appear in stored media. Research on Bayesian applications has led to many commercial applications, including handwriting and fingerprint analysis and facial-recognition software.

MBC employs the Intelligent Data Operating Layer (IDOL) to search unstructured data. IDOL parses very large data sets, using a combination of speech analysis, text comprehension, and "implistic query" to establish relationships among specific files that reside in unstructured data resources. Discovery of these relationships enables MBC to make inferences about implied, but perhaps "hidden," meaning even as modes of conversation and references to ideas shift and evolve in real time. Therefore, IDOL not only mines data but also finds associations among data at much faster rates than human beings can. Lynch's theoretical work has led to success in the marketplace: Autonomy, founded in 1996, is now a FTSE 100 company with a $7-billion capitalization and a long roster of blue-chip global firms.

BUSINESS LAW PROVIDES PROOF OF CONCEPT

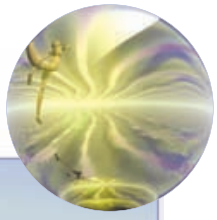
The most concrete example of MBC's impact is found in the legal profession. The legal world often plays the role of canary in a coal mine when it comes to disruptive technologies and their impact on knowledge workers. Long viewed by fledging attorneys as a safe, profitable, and guaranteed ticket to the upper class, business law firms are currently struggling. The recession plays a role in this unsettled period, but emerging technologies also have a large effect. Law firms earn revenue from billable hours, and so the number of hours a firm can bill provides a crucial revenue stream. Various aspects of legal work are slow and onerous and thus generate large numbers of billable hours. The best example is the discovery process, which involves attorney review of documents to find clues and references that can strengthen legal arguments. This is exactly the kind of labor that MBC can transform.

MBC analyses can reduce the number of billable hours charged for discovery by a dramatic margin, potentially saving corporate clients millions of dollars. As a result, firms such as Palo Alto, Calif.-based BlackStone Discovery are deploying MBC techniques. BlackStone provides e-discovery, sifting through millions of documents and producing high-quality results at a fraction of the cost of manual review. "We are living in an era when data of all types are connected, and these connections provide vital clues during discovery," John P. Kelly, president and CEO of BlackStone Discovery, told me. "I like to say that data 'are talking to each other in the third person.' If we can track this 'dialogue,' we can uncover associated documents, discussion threads and other references that otherwise might be missed altogether."

IMPACT ON KNOWLEDGE WORK

There are some clear value points to the MBC search technique. Most professionals, from accountants to engineers, devote a lot of time to research and discovery. MBC provides the means to extract fresh value from knowledge that was used once and then filed away—producing a new stream of value from old information. For example, engineers working on structural designs can employ MBC to search electronic mail, diagrams, and reports from the sum total of the firm's records. The results of experimentation on unrelated projects could help engineers find kernels of wisdom that have already been developed and apply them in new designs.

This engineering example implies that even though MBC can reduce the amount of work that is done in some steps of a large project, it may also produce strategic information that requires review and analysis. Therefore, despite *The New York Times*' focus on MBC's ability to replace highly paid lawyers, the software may also become another new tool instead of a complete replacement for human analysis. That may be cold comfort for knowledge workers who, once again, may need to retool how they do their jobs.



If MBC has thrown a challenge at the feet of attorneys, it symbolizes how technology now reaches into even the most hidebound of professions.

Just the same, the legal example of e-discovery bears a closer look. Can attorneys be replaced by automated routines in every instance, including the discovery process? The most accurate answer is, *in some, but not all instances*. The discovery process itself is only one aspect of litigation between opposing parties, and history is full of examples where fortune hung by a thread, a single document, or an elegantly managed courtroom drama.

The 1986 battle over Johnson & Johnson scion J. Seward Johnson's vast estate is a case in point. The case ran at high-octane levels, with a new and much younger wife at odds with family members and charitable foundations over the role of the new Mrs. Johnson in her husband's will. A single handwritten page was discovered and entered into the record late in the litigation process, after probate had been completed. It was found by a lowly legal assistant, and it provided (contested) evidence that Mr. Johnson had given his new wife broad powers in deciding estate matters. Would the document have been spotted by MBC techniques? Perhaps, but it would have had to be scanned for MBC techniques to discover it.

Moreover, the vast enterprise of discovery in the Johnson case was eclipsed by courtroom maneuvering that made the case a sensation in its day. This example demonstrates that effective legal work is the result of many activities and that MBC might act as a powerful, timesaving tool. BlackStone Discovery's Kelly concurred when queried about the Johnson estate example. "A handwritten note would have needed to exist somewhere within a repository, it's true. However, retrievable documents might *refer* to that note, and that would tip lawyers off to its existence. It's these kinds of references that make finding a single, all-important document more possible."

THREE AREAS TO WATCH

MBC is a powerful search technology, conceptualized as a means for extracting value from very large document repositories at the enterprise level. Unsurprisingly, this technology has only begun to be tested against new challenges. Indeed, machine-assisted analysis of language patterns that can be deployed across any type of media is likely to become an essential element in the information landscape. It could certainly become a helpful tool in the arsenal of information management techniques already deployed in the library and information services world.

If Web 2.0 and the app marketplace are any guides, we can expect some surprises, perhaps in the area of searching the dynamic web. In the meantime, here are three areas to watch as meaning-based computing advances beyond the commercial sphere and more deeply into our professional lives.

Taxonomy: It will still matter. The contemporary web is a matrix of artifacts and living documents that have their own "social life," as Xerox Palo Alto Research Center's John Seely Brown famously said in the book, *The Social Life of Information*, which he co-authored with Paul Duguid. HTML5 and the semantic web will add embedded functionalities that further support the development of taxonomies, ontologies, and crowdsourcing techniques to tag useful resources. Meaning-based computing will not supplant these features, although it will likely introduce powerful new ways to look inside document wrappers and parse the evolving meaning of everyday language. If so, then MBC presents digital content managers with another new opportunity to improve access and retrieval.

Repositories may gain ground. In 2009, ITHAKA S+R's respected faculty survey reported that repositories are of less interest to research faculty than we might hope. Better language analysis may give digital repositories a much-needed boost and increase their relevance for scholars. MBC is based on the awareness that context influences meaning and evolves over time, creating new ways to state facts and new ways to hide them. Historians will be the first to tell you that uncovering and tracing meaning in texts is half the battle in scholarship. Search tools that enable scholars to review language relevancy within large digital repositories clearly will have utility, much in the way that geospatial systems have formed the foundation for data mashups using government information.

Reference: An evolutionary "context." In the film *Desk Set*, Spencer Tracy and Katharine Hepburn do battle over the question-answering power of computers and survive the duel. More recently, IBM's Watson gained fame as a competitor

Elements of Meaning-Based Computing

Structured Data
Includes data residing in controlled formats such as databases, repositories, spreadsheets, etc.

Unstructured Data
Includes data residing in myriad locations, including electronic mail, audio files, podcasts, and loosely organized personal computers.

The Intelligent Data Operating Layer (IDOL)
Searches and processes all types of data, including both structured data (controlled input and output, e.g., databases) and unstructured data such as email, audio, and unorganized documents; penetrates "silos" of unstructured data and cross references it with structured data.

Speech Recognition
Technology is already employed widely; MBC improves search results by looking for evolutionary meaning in the way we use words.

Text Comprehension
MBC looks for similar words in every context and evaluates the probability of associated meanings across very large data sets.

Sentiment Analysis
Uses IDOL to track how people used words and uncovers changing patterns in meaning. For example, an email message that says "I'll call you" may imply a need for secrecy or privacy.

Implistic Query
Employees Bayesian probability principles to discover associated meanings in real time; may lead to functionalities such as instant access to media that share context with current work.

Principle Applications

Extract value from unstructured data
Risk analysis of data repositories
Tool for improving strategic planning

on the game show *Jeopardy!*, triggering fresh debate about the balance of power between people and machines. But as contextual tools such as MBC merge into collaborative workspaces, reference work stands to gain ground. Collaboration is the driving force in higher education, with multidisciplinary curricula, online teaching portals, and online learning spaces leading the way. MBC is just as likely to make reference providers and other sharp team players shine in the collaborative workplace as it is bury them in the dustbin of history.

Our field has thrived over time by integrating new technologies within the context of our core competencies.

CHALLENGES OF MBC

New technologies frequently deliver unforeseen results when they are deployed in the marketplace. The information professions—including not only librarians but also every kind of information-handling worker—have faced constant waves of disruptive technology for the past 35 years. Audacious new technologies appear and seem to threaten our obsolescence with such regularity that our dialogue about the future now embraces the challenge of responding to innovation, instead of flinching at the prospect. If MBC has thrown a challenge at the feet of attorneys, it symbolizes how technology now reaches into even the most hidebound of professions. But as BlackStone's Kelly suggests, MBC-generated knowledge still needs to be interpreted and integrated. Therefore, even as MBC extracts new value from rich media, we can expect to see an ongoing need for human review that goes hand in hand with timesavings.

Our field has thrived over time by integrating new technologies within the context of our core competencies—collection development, interpretation, preservation, and counsel. The best strategy is to analyze MBC and its potential in the context of our core competencies, which provide a unique vision of how the information society can function. That is an evolving context in itself, and as such, it is a solid foundation for integrating MBC into the rich media we already manage.

Terence K. Huwe *(thuwe@library.berkeley.edu) is director of Library and Information Resources at the Institute for Research on Labor and Employment, at the University of California–Berkeley.*

Comments? Contact the editor (marydee@xmission.com).