

subscription to view the full text. For the journals, links that are labeled “Science Direct” indicate an Elsevier full-text article.) While viewing the window with the individual records, we could also click on “Profile” to see a summary of all of the information that illumin8 has collected on that entity.

There were some other search features available to us as well. We could change the results’ default to rank the results by relevancy or total count, do a “search within results,” and export our results to Excel. And at any time, we could click on the Search Guide for very helpful and informative search tips and advice.

Pricing

Buzzanga told us that pricing is by annual subscription, tied to the number of users. When we asked for a rough idea of cost, he told us, “We are not releasing pricing as a matter of policy.” However, a NewsBreak article written by Barbara Quint on Feb. 28, 2008 (<http://newsbreaks.infotoday.com/nbReader.asp?ArticleId=41084>), quoted Elsevier’s vice president for product development in engineering and technology, Rafael Sidi, that the base price for a Fortune 1000 company starts at \$100,000 a year and then builds based on the number of users. A spokesperson for illumin8 told us that current prices, though, are lower than that.

Evaluation

So what did we think of illumin8? Did it deliver on its promises? Did it locate what we could legitimately call benefits to the technologies we entered and approaches to the problems? Did it also identify relevant companies and products? And did it offer the landscape approach for our subject queries? Even more to the point, did it work? That is, were the results relevant, substantive, and did they answer our questions?

Overall, yes, and we were very impressed with illumin8, for the following reasons:

- Most importantly, it did work. Results that were categorized as “benefits”

We liked that we were able, immediately and at a glance, to get a sense of the most important elements, issues, people, etc., for the topic we entered.

largely really were benefits, “approaches” were approaches, and so on. The NLP was impressively smart. (Buzzanga explained that illumin8 “searches over pre-selected text snippets that the NLP engine has selected based on meaningful research-relevant relationships ... like: X solves Y, X is a benefit of Y, X sells Y, X is an approach to Y, X invented Y, etc.”) So when we search, we are “searching over” these semantically extracted text snippets.

For example, when we keyed in the problem of “high cholesterol,” the “approaches” that were returned were, in order: Exercise, Screening, Lose Weight, Drug Therapy, Statin Therapy, Bariatric Surgery, Education, and Lower High Blood Pressure.

- We liked that we were able, immediately and at a glance, to get a sense of the most important elements, issues, people, etc., for the topic we entered. In nearly all cases where we searched on a subject we knew well, more than 80% of what we viewed were entities we would agree were the most important ones related to that subject.

- illumin8 offers a new way to search and another legitimate alternative to keyword searching and its refinements (e.g., link analysis, social search) by combining an effective NLP software with a clustering results page against a specified body of high-quality content.

- The search experience itself was quite pleasing: intuitive, fast, well-designed, easy to use, and with some excellent help screens.

On the down side, we also discovered the following:

- There were obvious errors with the NLP software. We found that “carbon

dioxide” was categorized as an approach to solve the problem of global warming, a bird was an approach for the technology of “wind power,” and so on. These kinds of errors go with the territory when it comes to using NLP, and, as we said, we were pleasantly surprised by how accurate it was. But you still have to deal with the obvious—and the more hazardous *not* so obvious—errors.

- We would like to have had more flexibility and options in our searches: e.g., the ability to limit to certain publications or authors.
- While we haven’t been able to get Elsevier to reveal its pricing, we think that what seems to be a premium subscription-based service should offer more extensive customer support and help options. Currently, there is no promise for quick phone support turnaround. And we hope that less expensive versions are made available for smaller firms—and soon!

Overall, we’d recommend illumin8 as an excellent choice to get an overview and initial review of the key issues surrounding a scientific, technological, or technical matter. And we’d use it for its unique ability to surface approaches and benefits. Because NLP is an imperfect technology, you’d still need to be careful and make some initial guesses on the significance and relationship of each entity. However, we found that the results and the reason for including certain items gradually becomes clearer by reading records. ■

illumin8 Shines a New Light on the Search Process

We know that plain old keyword searching usually works pretty darn well. Whether it’s searching a traditional online database, going to Google, using a blog search engine, or searching the contents of an individual website, the keyword approach to finding information is extremely simple and still largely effective. But we also know that keyword searching has its limitations. This method for finding what we need still represents a rather crude approach for matching up what we know we want to find with the actual *meaning* of the documents, sites, articles, or other sets of information we are sifting through.

For many years, database vendors, analysts, publishers, library technology firms, search engines, and others who have an interest in improving the search process have embarked on their own searches for a better method to identify the most relevant and meaningful information. Perhaps the best known of these approaches, and the one that’s been talked about the most for a very long time, has been the goal of building a smart or “semantic web.” A seminal piece on this was written by Timothy Berners-Lee titled “The Semantic Web,” published in *Scientific American* in 2001 at www.sciam.com/article.cfm?id=the-semantic-web.

But creating a smarter web where the pages have embedded meaning has proven to be difficult, if not impossible, to achieve. Other approaches have tried to improve search relevance and impart more meaning. These newer methods have included, for instance, simple

page clustering, link analysis, and disambiguating terms, which is a kind of semantic search “light.” That disambiguation has been done at least partially successfully by a few search engines such as hakia and Cluuz (see our review of these “alternative” search engines in the December 2008 issue

Creating a smarter web where the pages have embedded meaning has proven to be difficult, if not impossible, to achieve.

of *The Information Advisor*). Most recently, there has been an attempt to introduce social-oriented solutions, such as collaborative filtering, having users add their own tags (folksonomies), and “social search” where results are based on what pages and sources your friends and colleagues have selected. All of these approaches have been introduced with the aim of better ensuring that searchers find what they really want to find.

One new and, we think, notable entry in this effort to improve on keyword searching is a product called illumin8 (www.illumin8.com), introduced early last year by Elsevier, the giant scientific, technical, and medical (STM) information publisher and information service. To create illumin8, Elsevier partnered with a technology firm called NetBase (previously known as Accele-

vation), which supplied Elsevier with the natural language processing (NLP) software. That NLP technology was developed at MIT Media Labs in 2003 by NetBase co-founders Jonathan Spier and Michael Osofsky.

We had a chance to get a full demo of illumin8 and then spent time trying it out ourselves. We present the results of our testing and our evaluation in this article.

What’s Inside illumin8

Before getting into the technology of how illumin8 works, let’s first describe the content. illumin8’s collection focuses on science, science-oriented business, and technical areas; it consists of articles from scholarly journals as well as selected webpages. Specifically, it includes the following:

- 36 million scientific abstracts from 15,000 peer-reviewed publications from 4,000 publishers in the STM fields: Joe Buzzanga, product manager of Elsevier Engineering and Technology, told us that the aim of the collection was to “cover all of the peer-reviewed literature in those fields” and that this represents the largest abstract database of this type of information in the world. New articles are added weekly to biweekly.
- 3 million full-text scientific articles from 1,800 journals published by Elsevier: New articles are added weekly to biweekly. Note that this collection is the same as Elsevier’s Science Direct database. Buzzanga told us that it is the equivalent of that database’s corporate edition,

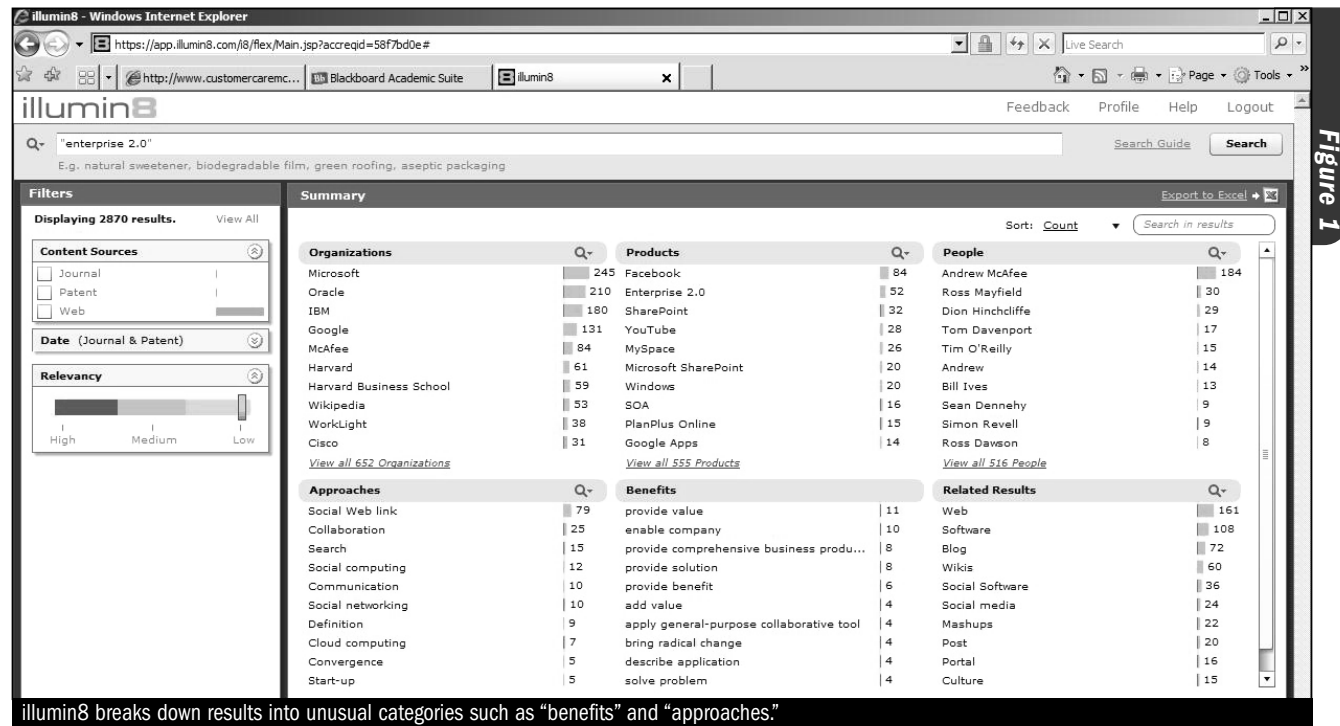


Figure 1

illumin8 breaks down results into unusual categories such as “benefits” and “approaches.”

with coverage dating from 1995.

- 22 million U.S. and global patents, from five worldwide patent offices.
- Approximately 8 billion English-language, scientific, and business-related webpages, blogs, and forums: To gather these pages, illumin8 uses its own web crawler, but it does not refresh pages frequently, so it should not be considered.

Unlike most search engines, illumin8 works best if the search words describe either a problem or a technology.

Note that illumin8’s collection strongly emphasizes science and focuses on R&D, as its broadest goal is to assist firms to do a better job in its innovation creation process. However, the collection is science broadly defined, and it encompasses technical, management, and R&D topics and issues relevant to many business research queries as well as to sci-tech searches.

It is important to note too that, unlike most other search engines, illu-

min8’s goal is not just to make searches more effective—though that is a key part of its mission. illumin8 was also developed to assist with a particular *type* of research question or problem: that is, to help the researcher (and organization) find the best or most frequently cited *approaches* to a problem, as well as locate the *benefits* of a certain product or technology. The layout and components of its results page (see Figure 1) also was designed to reflect illumin8’s larger philosophy of information and research. As Elsevier put it: “illumin8 is a search engine designed to find solutions and create *technology landscapes* [reflecting] *the idea that every piece of information exists within a web of relationships and, when organized, the relationships can be viewed as a map or terrain.*”

This focus on surfacing *approaches* and *benefits* is noteworthy, as it is conceptual queries such as those that have always been much harder for standard search engines and basic keyword searches to handle. If, in fact, illumin8 has made headway in providing more relevant and useful answers for these kinds of searches, it would represent a very notable search innovation.

As with all efforts to build a new and supposedly better information trap, the proof is in the searching. So we spent

some time doing our own searches on illumin8, which we detail here.

Searching illumin8

The way we began searching illumin8, as we do when using nearly any other search engine, was just to enter a word or phrase into the site’s search box. However, because of the distinctive nature and approach of illumin8, the searcher needs to think a little more carefully as to what words or concepts would be best. Unlike most search engines, where one might casually enter whatever words or phrases that came to mind to describe a subject, illumin8 works best if the search words describe either a *problem* or a *technology*. In this way, the results will surface solutions to the problem and benefits of the technology. (illumin8 also retrieves other kinds of results, such as associated people, products, organizations, etc., which we will describe later.) An example of a good search term for illumin8 might be “global warming” in order to find solutions to that problem, or “wind power” to find associated benefits of that technology. These distinctive search strategies are outlined nicely in illumin8’s very helpful search guide with tips like these:

- Base your search query on questions, not potential answers.
- When you have a problem, the

temptation is to search for an answer that you already know about. With illumin8, you can break free of preconceived ideas and find results that you might not suspect.

- Describe the problem when you construct your search query. If you want to find a product that helps relieve dry skin, don’t run queries such as “dry skin products” or “skin lotions.” Instead, type in “dry skin,” add a few synonyms such as “chapped,” and then run the search.

As for illumin8’s search protocols, entering multiple words defaults to AND, with a relevancy ranking formula applied if both terms are not found; quotations can be used for phrases; and there is support for Boolean AND, OR, and “E” for exclude. An advanced search page (called Search Guide) allows for three field searches as well: subject (the default search), technology (in which the system treats the query term as a technology term), and attribute (benefit, problem, or characteristic). Buzzanga says that the attribute field can be useful if “the user is very certain of their search: for example, the query ‘wireless network congestion’ is pretty clearly a problem statement and not a technology.” The advanced search page also suggested synonyms for subjects we entered.

Note that there is no option for limiting a search to specific publications, websites, or other subsets of illumin8’s full information collection.

After entering our search terms, illumin8 took about 5–10 seconds to create

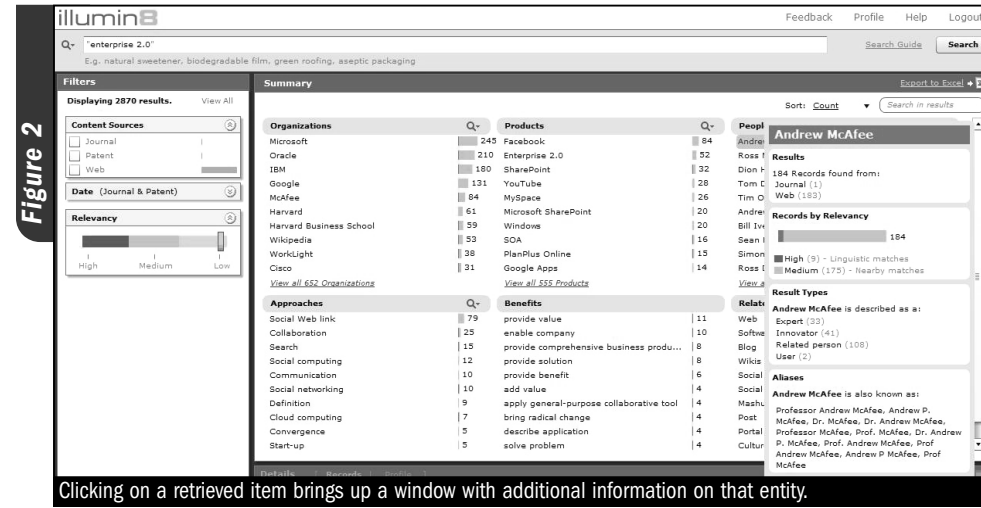


Figure 2

Clicking on a retrieved item brings up a window with additional information on that entity.

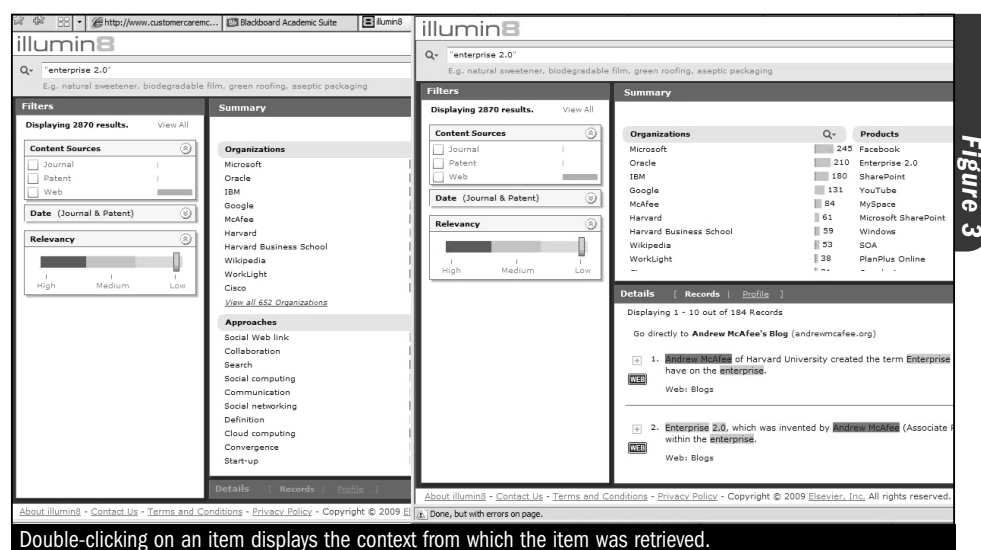


Figure 3

Double-clicking on an item displays the context from which the item was retrieved.

its very interesting results page (see Figure 1). The results page is broken down into two areas: a narrow vertical bar on the left and a much larger window on the right. These contain the following elements:

Left-hand bar:

- Total number of results from the search.
- Content Sources: Journal, Patent, Web. A green bar indicates roughly what percentage of the results were retrieved from each of those collections.
- Date (for Journals and Patents only). A green bar next to each year indicates roughly how many items were associated with a particular year.
- Relevancy: High, Medium, Low. This code is used to evaluate the relevancy of the matches on the right-hand portion of the page. The

darkest portion of the bar indicates a strong NLP connection has been found; middle green is less strong; and light green is weakest.

Note that all of these elements serve as filters for refining the same search. For example, we could choose to run our search again and view only journal, patent, and/or web results and filter the results by year or by level of relevancy.

The actual results are displayed on the right side of the screen, and all results are listed under one of the following categories: Organizations, Products, People, Approaches, Benefits, or Related Results. Under each of these headings, a particular organization, product, person, term, or other element is listed, with the top elements on the list having the highest count or relevancy. Mousing over any individual entity (see Figure 2) provides additional information on that item: a breakdown of how many records were found from which collection (patent, journal, web), a breakdown of how many of the records were classified as high or medium relevance, and a breakdown of the subcategory TYPE of organization, product, person, approach, benefit, or related result. For example, a person could be further described as an expert, an innovator, a “related person,” and so on.

Double-clicking on any of those entities opens up a window at the bottom of the page (see Figure 3). This displays the result within the original context in which it was found. We could also click on a link to go to that original source. (If the source was a subscription journal, the user’s institution must have a