Source: Knovel

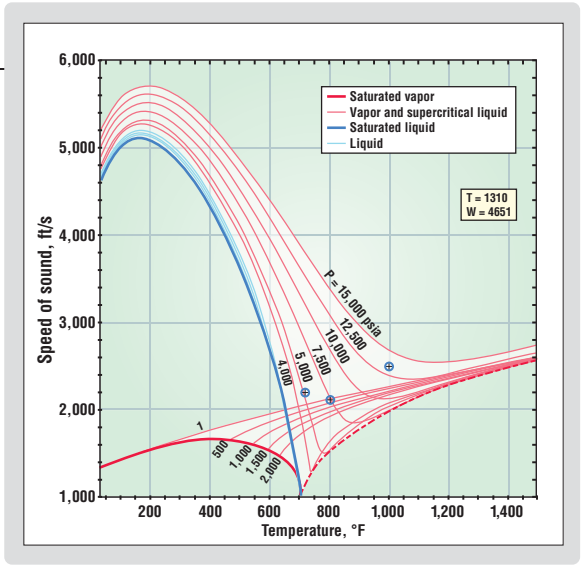

FIGURE 1. Information resources, such as Knovel, allow users to manipulate data of the kind that at one time were static, and adapt that data to their own engineering problem

INFORMATION GETS DYNAMIC

Enhanced chemical information tools help CPI engineers manage ever-increasing volume and tightening work demands

As engineers in the chemical process industries (CPI) are pressed harder to meet deadlines in the workplace, the amount of information available to them in their field is growing at an exponential pace. To address the situation, chemistry- and engineering-oriented information resources are developing tools and systems that offer opportunities to process information more efficiently and interact with available information more effectively.

Among the tools becoming available and becoming more widely used are those for enhanced search capabilities, those for refining search results, and those enabling direct interaction with, and sharing of, data. To a greater extent than ever before, the capabilities of information resources for CPI workers emphasize how the information is used once it is located. Tools now make it easier to adapt data directly to apply it to a specific problem relevant to the searcher, as well as make it easier to share information with coworkers.

As information resource platforms evolve, changes in the way engineers in the CPI work are occurring in parallel, and the changing needs of engineers are having a significant influence on what capabilities are offered by information resources. Added into an already complex interplay is the maturation of wider technological trends, such as social networking, community forums and web-enabled mobile devices. The result of the convergence seems to be an ever-richer environment for chemical engineers to access and use available chemistry-related information. The improved set of tools can help address the increased pressure felt by CPI professionals to conduct comprehensive information-based research in a timely fashion.

Many factors drive demand

Because of its global nature and strict time demands, modern engineering work places a heavier burden on information resources. To a large extent, today's CPI is characterized by geographically dispersed corporate structures and leaner, less specialized engineering teams. The result is that interdisciplinary know-how becomes a premium, and training and knowledge transfer become more important. In addition, the ability to effectively communicate, collaborate and share information across continents and time zones becomes critical to success. Meanwhile, engineering teams are under greater time constraints than ever before, and the pressure to rapidly find correct answers on project-related questions is equally high.

The situation is rendering information resources indispensable, and the developers and curators of such products are answering with a fundamental focus on quality. Although centralized databases of computerized chemical information are important, and the ability to parse a torrent of available information is equally so, access to them means little without quality.

"The integrity of the data contained in a database is of utmost importance to engineers," says Shandon Quinn, director of market development for chemicals at global scientific publisher Elsevier Inc. (Amsterdam; www.elsevierchem.com), owner of the information platform Engineering Village (EV; www.engineeringvillage2.com). The subscription-based service hosts a number databases, including bibliographic databases, European and U.S. patent databases, news and others, making it a "one-stop shop" for several different types of useful content. One key area of emphasis for EV has been information quality. The EV team operates a content management system in which information sources are continually scrutinized to ensure that the highest-quality information is included, Quinn comments.

Organizers of other resources place information quality as a high priority as well, and have set up systems of their own to select the highest quality data. For example, Knovel (New York; www.knovel.com), operator of a Web-based application of the same name that integrates technical information with analytical and search tools to aid in problem solving, maintains an editorial advisory board to pre-validate information for its customers, and to guide content additions. The subscription-based content aggregator and search engine allows users to gain access to engineering reference content sourced from professional societies, publishers and authors. Knovel offers access to a mosaic of different types of information, from material data sheets and technical references to interactive

MAKING CHEMISTRY-SPECIFIC USE OF THE SEMANTIC WEB

In coming years, it is likely that the volume of information available through the Internet will continue to grow unabated. Because of this, more powerful tools will be required for processing this information and extracting useful items for a given searcher. Future Internet-based searching will increasingly have access to artificial intelligence systems that are capable of absorbing the meaning from textual language, rather than just identifying keywords.

The Semantic Web is a term that describes what many see as a part of the Internet's next evolutionary step. The term refers to a virtual data network that enables machines to understand the meanings of language in Web-based information. The Semantic Web extends the network of human-readable Web pages by adding machine-readable metadata (descriptors of the content and context of data files) about Web pages and how they are related to each other. This enables automated agents to access the Web more intelligently and perform tasks on behalf of human users.

Much effort is being devoted to emerging tools across many fronts, and chemistry is among them. Peter Murray-Rust, professor at the University of Cambridge (Cambridge, U.K.; www.ch.cam.ac.uk) is one of a cadre of researchers interested in the future of chemical informatics. Murray-Rust has developed Oscar (Open-Source Chemistry Analysis Routines), a software tool that is designed for the semantic annotation of chemistry documents and papers. The software package allows chemistry-specific parsing of documents — identifying chemical names, formulas, acronyms, as well as ontology terms and chemical data, such as spectra, and other experimental data.

Murray-Rust treats chemistry as a language that is communicated not only through natural language, but through formulas, equations and graphics. Oscar can recognize names in text and is able to link them to their meaning through their ontological relationships. Key to the software is its ability to process natural language and discern the context in which chemistry-related words are used. The tool, which can be downloaded and tested for free, could save time and effort in research and data gathering by better filtering out only the most relevant hits, while ensuring that nothing with an equivalent, but different name, gets missed

The idea is to automate the extraction of knowledge from text for later use by machines, or to give human searchers additional information about the text, such as whether the author of a given paper agrees with a previous publication, or is criticizing it. Murray-Rust and colleagues are trying to combine "shallow," fast-processing approaches that can identify parts of speech (deciding that a word is used as a verb or a noun in a sentence), with "deeper" systems that require more processing time, but that can decipher more meaning, such as using context to decide whether a phrase is meant literally or idiomatically. The combination would allow the processing-heavy system to be used only on text that has been identified as interesting by a faster-running shallow system. Murray-Rust is working on trying to incorporate a method to automatically discover information about the meaning of chemistry terms into his chemistry search tool. ❑

tables and conference proceedings.

Chemical Abstracts Service (CAS; Columbus, Ohio; www.cas.org), a division of the American Chemical Society (ACS; Washington, D.C.; www.acs.org), also devotes considerable attention to information quality. The world's largest collection of chemical information, CAS employs a dedicated editorial staff to curate the content found in CAS databases to ensure its reliability and quality. "The volume of available information is huge, and part of that is of poor quality," says Kirk Schwall, CAS director of SciFinder product development. "Our challenge is to make sure that we are offering the most reliable peer-reviewed information, and nothing else."

Aside from ensuring data quality, providers of information resources are also focused on several other capabilities in the tools that they offer. One of these capabilities is rendering data and information interactive, to enhance their value for users. Another capability is optimizing searching specifically for engineers to make research more efficient and productive. Other aspects of providing content that information resources view as important are extending a tool's reach to mobile data platforms, such as Web-enabled mobile devices, and continuing to make additional content available in areas of current interest to chemical engineers. Knovel founder and chemical engineer Sasha Gurke notes that to meet demand from users, his company has made additional content available in areas such as nanotechnology, process automation and control, alternative energy and sustainability.

Optimizing search for engineers

A major push to improve information tools in chemistry-related fields has been to optimize search capabilities specifically for the needs of chemical engineers and research and development personnel. "Google only retrieves publicly available information, so it misses some premium providers of information" such as professional groups, publishers and authors, explains Knovel's Gurke. Also, chemical engineers require databases with chemistry-specific search capabilities.

"Considering the time constraints experienced by most working engineers, the precision of search is critical," adds Elsevier's Quinn, "but there remains a pressing need to produce comprehensive results at the same time, so that engineers are confident that they are not missing anything important to their project."

Elsevier's Engineering Village is an example of an information resource that has adopted this approach. The Web-based information services platform hosts a number of interoperable databases specifically targeted toward engineering disciplines. EV concentrates considerable effort on establishing and maintaining "controlled vocabularies" that standardize engineering terminology so that indexing is more efficient, and information is better organized for users. For example, some sources might use the term "biopolymers," while others, meaning the same thing, might use "bioplastics." EV's controlled vocabularies index positions the terms in a way that allows searches return as many relevant records as possible.

"Today's engineering work is more interdisciplinary than it has been in the past," comments Elsevier's Quinn. "So there is a greater need to understand terminology across a number of fields, and that's a big reason why the standardized indexing is so useful."

Other features aimed at improving searches for chemistry provide alter-

natives to keyword searching, which may not always be the most useful method for searching in chemical engineering and other chemistry-heavy disciplines. In many cases, it may be more successful to search by chemical structure, or by reaction pathway, for example. On Knovel's platform, users can search numerically, and will soon have the ability to search for mathematical equations. CAS has introduced functionality to its platform that allows users to search for Markush structures (generalized formulas for a set of related compounds). It will allow searchers to identify a family of chemical compounds that share a common chemical backbone, for example, but may differ by a side chain. EV also has a search function that goes beyond keywords, allowing chemical structures and reactions to be search subjects.

Faceting of search results is another key consideration. Faceting refers to the dynamic clustering of search results into categories that allow users to drill into specific areas of interest while easily excluding those areas that are not relevant to a particular search. This month, Knovel will launch a newly revised platform with faceted search results as a major feature. It will allow users to view and filter search results by content type. EV also focuses on faceting as a quick and useful way to refine search results into a manageable set of highly relevant results.

Allowing interactivity

Aside from making it easier to locate and organize items of interest, information resource-tool suppliers are trying hard to give the search results more value by introducing the ability to interact directly with the data and information. Summarizing the philosophy, Knovel director of marketing Diana Bittern says, "We give Knovel users the ability to connect to information in a way that's most meaningful to them, and then interact with it to create additional value for them."

Knovel enlivens reference data by introducing interactivity and allowing engineers to manipulate data — such as graphs, equations and tables — and incorporate them into their work though programs like Microsoft Excel or PTC's Mathcad. For example, users can interpolate a plot, or adapt an equation for use in their own situation.

Knovel's Gurke says that recently there has been a need to combine data with calculations software. For example, after finding a useful calculation example, Knovel users can design directly on the platform with Knovel Math software, allowing them to obtain a validated solution right away.

Further, Knovel gives users personalized access to a shared workplace, where they can work interactively and dynamically with content that, in the past, would have been static in a

book or other printed publication, explains Bittern.

CAS recently launched a new tool that also facilitates interactive work. Since 1995, CAS has offered chemical researchers and industry professionals SciFinder, a single-source platform for a wide range of chemistry-related databases. Now, a newly launched feature, called SciPlanner, has been added. SciPlanner is an interactive workspace that represents a new way for scientists and engineers to more quickly identify synthesis options for designing reaction pathways and approaches. The feature also helps to organize results and manage data. SciPlanner's interactivity is a key feature, with a live, virtual whiteboard — a kind of "smartboard" system — that enables work to be subsequently shared with colleagues.

Among SciFinder's features are the ability to sort reference search results by citation count, to identify influential authors by the level of citation, and the ability to copy and paste drawn chemical structures directly into the SciFinder structure-drawing editor to save time. A version of SciFinder that is compatible with Web-enabled mobile devices is also available.

Peer connections

Mobile applications are only one of a host of developments that are poised to become more widely used in the CPI space, as they are in other sectors. Examples are crowd-sourcing, social networking and community-editable Web pages (wikis), which are the types of elements supported by community forums. A prominent one for the chemical engineering community is CheResources.com. The platform offers tools of a more organic kind, such as those for when chemical engineers look for direct interaction with peers, rather than with databases. CheResources.com has built a strong following with a core group of experienced engineers who relate experiences and advice.

"We offer high-quality advice from experienced engineers on real problems," says Chris Haslego, founder and community manager of the site. The Website has been transitioning from a more static, article-driven site in the past to a more community-driven format, where blogs, question-and-answers, and postings make up most of the new activity. "We're moving to a Website that is much more dynamic," says Haslego.

"Our next challenge revolves around downloads and mobile applications," he adds. The site has recently introduced a download section, where members can download and use spreadsheets that have been developed by others, saving time for engineers, Haslego explains further. Interactivity is the key, because they can adapt the spreadsheets to their own problems. ■

Scott Jenkins