Content Management Systems – The Next Generation

A new generation of content management systems addresses four key areas that were hindering search and retrieval in the pharmaceutical sector – accuracy, security, scalability and multilingual functionality.

By Tim Ames, International Business Development Manager at Convera Technologies

In the 1990s the Internet revolution accelerated innovation, and businesses started to work in 'Internet time' – adopting a 'build it and they will come' attitude. The pressure was on to bring products to market more quickly to keep ahead of the competition. As a result, traditional R&D practices were put on the back burner; the fact that R&D was the idea engine, the generator of innovation and the key to success or failure, was largely ignored. This lack of emphasis on R&D was reflected in the number of companies and technologies that disappeared as fast as they had appeared. Now, industry is experiencing a snap-back to traditional paradigms, with an increased focus on R&D to ensure that market demands are met and companies can effectively predict what the marketplace desires – whilst continuing to innovate.

One industry that could never afford to change its R&D practices so radically was the pharmaceutical industry. Despite many dot.com health enterprises collapsing, or experiencing difficulties as the business models they pursued proved unsustainable, traditional health care business models proved robust. For these companies, the Internet and technology simply increased R&D momentum – the key being the management of a company's intellectual property. By embracing Content Management (CM) as an efficient method to access, analyse and share information, the pharmaceutical industry has been successful at innovating in a highly competitive market; by reducing the time, money and resources spent during the R&D process, 'time-to-market' has been significantly reduced.

Within the pharmaceutical industry, the need to be first to market for any given drug is paramount. Drugs can take many years to develop, test and achieve government sign-off before marketing and sales activities can commence. By using CM to help streamline the process, even relatively small gains in 'time-to-market' of 1-2% translate into a drug appearing weeks or even months ahead of normal schedules.

Any pharmaceutical company not taking a serious interest in CM tools and policies is tantamount to committing commercial suicide. Next to laboratory work, information gathering and analysis are the most time-consuming and mission-critical activities in pharmaceutical R&D. Scientists need to have fast and easy access to the latest findings on drug development in order to survive in an increasingly competitive field. However, these findings exist in many formats and are distributed across many departments, both internally and externally. As a result, the research and analysis of such information has always been a long and painstaking process, and traditional CM processes are often hindered by the complex sensitivities that are intrinsic in the pharmaceutical sector.

Out of the ashes of the Internet revolution, a new breed of CM system has emerged that effectively overcomes the restrictions faced by the more traditional models by incorporating extremely intelligent, pliable search and retrieval tools. Providers of next-generation CM tools have identified and addressed four key areas that were hindering traditional search and retrieval systems used within the pharmaceutical industry – accuracy, security, scalability and the ability to search and retrieve relevant documents across a multitude of languages. The IT services company, Ovum Consulting, defines next-generation search as “the technologies and products that are bringing new levels of intelligence, order and personalisation to the search process” (1). By implementing next-generation search, CM tools are tackling these four key areas and taking control of the vast amounts of information generated by the networked world.
**ACCURACY**

Most knowledge-workers want to be able to search in much the same way as they speak, but too often they end up with queries that are either unfocused and return thousands of hits, or are too narrow and overlook the important results. Building on an already complex process, next-generation CM tools have evolved to enable ‘natural language querying’; this frees up users to search in the way that they think, in return for a very specific results list with the most accurate documents at the top. It does this through a combination of concept- and pattern-searching which results in fast, accurate and highly relevant results. People naturally think in terms of concepts, not keywords, and their searches are often exploratory in nature. They don’t want to spend time figuring out the right keywords, but would rather enter a concept that translates a request into a complete search for all relevant and related concepts.

Combined with pattern-searching to identify misspelled words, this approach has been applied to the pharmaceutical sector and concepts for a whole new language – the complex language of pharmaceuticals – have been created. By incorporating semantic networks to automatically augment the query through pharmaceutical-specific concepts and patterns, any terms and associated drug names that may have slipped the mind of a researcher, or that may have been spelt incorrectly, will still be included in the results returned. Furthermore, a number of next-generation CM providers offer ‘plug-ins’ enabling R&D teams to further drill down the search criteria of a specialised field, or to incorporate unique department-specific vocabulary into the system. An example could be for a team developing research on a particular skin complaint to ‘plug in’ concepts that are exclusive to dermatological agents.

Without pharmaceutical semantic networks and associated ‘plug-ins’, it is the person searching for the information that determines the quality of the hit-list. For example, it is rare that everyone in a research group would include all of the drug names associated with anti-psychotic drugs when conducting a search on this subject. With an extensive list of associated drugs available, the right semantic network would include them all in the results – regardless of whether they were within the search criteria or spelt incorrectly.

According to research by IDC (2): “Timely access to critical information separates the winners from the losers in today’s information economy. Without this level of accuracy and therefore timely access to data, many real-world experiments would have to be reconstructed many times over, simply because researchers cannot easily find the results of previous experiments in the corporate information stores”. To put this into perspective, leading industry analysts have calculated that an enterprise employing 1,000 knowledge-workers would waste approximately £32,000 per week – or nearly £1.6 million per year – due to an inability to locate and retrieve information that is known to be ‘somewhere’. With next-generation CM systems, potentially valuable information is not lost from the hit-list, and time will be saved later as the research becomes more advanced.

**SECURITY**

The issue of security is one of the biggest hurdles the pharmaceutical industry has had to overcome in terms of sharing information and knowledge. With so many contractors and external sources working within the corporate pharmaceutical sector, it is not surprising that paranoia prevents legitimate teams from sharing potentially useful information between projects – the preferred option seemingly being to ‘lock down’ data and keep it secret.

With R&D teams often working across several continents, it is essential that companies facilitate collaborative development through the use of CM systems. Unfortunately, many companies are counter-productive when it comes to the adoption of such technologies; as a result of their focus on ‘keeping the bad guys out’, the needs of legitimate users from both inside and outside of the organisation are ultimately negated. On the other hand, some companies have implemented one or more individual security measures, believing themselves to be secure; in reality, however, the only way to create a truly secure environment is to implement a cohesive, holistic security policy.

Driven by such a demand, next-generation CM systems communicate and constantly update access privilege rights for all users in an automated and easy-to-manage way, offering uncompromising asset security across the entire information infrastructure. This is only made possible by ensuring that the search application recognises and enforces the security privileges of the organisation, such that users may only retrieve information that they are authorised to see. Next-generation CM models have incorporated a number of security functions to enable this:

- **Authentication**: The CM system provides users with a valid user name and password, as well as generating a unique session key that verifies the authenticity of every single transaction.
- **Encryption**: In secure mode, all communications (passwords, searches and search results) are encrypted to ensure that – even if intercepted by an unauthorised party – access to data will not be possible.
**Distributed security:** By permitting the distribution of processes across multiple systems for maximum scalability, a single log-on across multiple servers is possible, wherever they are located. This provides a distributed security capability that enables the CM system to share group information with other remote, trusted servers. For access to remote libraries, the user only has to log on once to the local CM system and – by proxy – is logged on to all other CM systems.

The resulting CM system creates a complete inventory of all available information assets, with associated access rights automatically restricting or expanding the search dependent on the level of security allocated to the user. Research indicates that 35–50% of the information available within an enterprise is not centrally indexed. The development of such an inventory is therefore essential to ‘free up’ important data. Used in conjunction with appropriate access rights, the resulting system allows users to search securely for relevant material in over 200 document types on a variety of file servers, groupware systems, relational databases, document management systems and web servers.

**SCALABILITY**

The pharmaceutical industry has undergone significant changes in recent years and, as a result, has seen increased M&A activity with several large companies merging to form ‘super-corporations’. Such mergers carry significant R&D implications and, if used effectively, can provide a good opportunity to leverage the combined knowledge of the merged companies. By determining whether inherited research projects are complementary to existing research, the newly merged company could deliver a drug to market more quickly than had previously been expected.

According to Gartner (3):

> “Discovering, developing and marketing a blockbuster drug is expensive and risky. From the time research begins to the time a drug is approved by the FDA... a drug company will spend £802 million... for every 5,000 compounds explored, only five make it to clinical trials. Of these five, only one will be approved for patient use. And less than one-third of marketed drugs achieve enough commercial success to recoup the company’s R&D investment.”

Following a merger, next-generation CM systems enable a corporation to integrate a whole new database of research into its search, retrieval and categorisation processes to identify similar projects undertaken by each of the individual companies; this can potentially result in significant time- and cost-savings. To be able to identify whether FDA approval has been declined, which compounds have made it to clinical trials and which unfinished projects can be utilised to bolster existing research is something that only scalable next-generation CM systems can offer.

**CROSS-LINGUAL CAPABILITY**

Pharmaceutical company operations are rarely constrained to one country and a huge amount of research material exists in many languages. The task of translating this content into multiple languages for the right projects and at the right time is prohibitive at best and, as a general rule, unworkable. However, next-generation CM systems make it possible ‘in real time’ to convert queries into multiple languages, and thus enable a user to know that a relevant document exists in another language.

Without next-generation CM systems, the complications involved in providing the right content are huge. How do you decide which documents should be translated and into which languages? How do you translate a huge database of documents in a variety of formats into multiple languages? How can you predict from which country an employee will request information? By incorporating next-generation CM systems, it is up to the user to decide, based on the results of a targeted search function, whether they need the information and whether they need it translated. The problem then becomes a ‘translate when needed’, rather than a ‘translate in case needed’ problem. The associated cost-savings are enormous.

**CONCLUSION**

By looking at these four areas, we can begin to understand how the intrinsic nature of the pharmaceutical industry has propelled CM software to a new level. By looking at the language problems associated with a global working culture, the sensitive nature of the R&D process and the complexity of industry-specific terminology, it became clear that CM systems needed to incorporate highly pliable and accurate search functionality. In doing so, next-generation CM technologies have evolved to empower innovation, reduce costs and deliver an increasingly shortened ‘time-to-market’ – offering pharmaceutical companies a significant return on the value of their associated investment.

The author can be contacted at tames@convera.co.uk

**References**