



systemware

White Paper

Deploying the “Find Engine”

An Automated, Indexed Approach Accelerates the Ability to Find and Use Key Enterprise Information

Systemware, Inc.
Dallas Parkway #1100
Addison, TX 75001
1-800-535-8681

www.systemware.com

A new generation of enterprise-class content management technology has emerged specifically to meet the content management needs of dynamic organizations.

Most corporations spend millions of dollars to generate business content – everything from application output to customer records, employee-generated documents, and transactional content. This information often resides in petabyte-size data repositories and represents one of the most valuable assets a business can possess.

But in all too many organizations locating and accessing key bits of that content is costly, time consuming, and frustrating. Companies have sought faster and more efficient ways to manage content across data warehouses and business units. Many use a business-oriented variation of a Web search engine, while others employ complex corporate taxonomy schemes in an effort to organize and use their vital business content. But those methods have delivered less-than-ideal results in the enterprise content environment.

Now a new generation of enterprise-class content management technology has emerged. Based on a multiple-level indexing approach and driven by advance categorization and presentation techniques, these new solutions are designed specifically to meet the content management needs of large and dynamic organizations.

Previous generation “search engines” were slower, labor-intensive, and tended to deliver large and unwieldy groupings of “results”. The next-generation “find engine” leverages advanced capabilities to more quickly and economically deliver only the precise and exact content needed in a given business situation.

In this Systemware white paper, we examine the importance of content search capabilities in the enterprise content management (ECM) environment, the capabilities of this new class of find engine technologies, and the benefits corporations can realize through faster and more precise content “finds”.

The Cost of Searching

Few business executives focus on the finer points of enterprise content management, nor should they. But the simple fact is: in today’s competitive environment, most companies rely heavily on their ability to quickly find and retrieve specific information from the growing corporate content stores, and then present that content in a usable format to employees, customers, and business partners.

Less-than-optimal content searches can put any organization at a real competitive disadvantage. Substandard search capabilities cost time and money while wasting both human and IT resources. Poor information flow often creates significant and costly lags between month-end closing and availability of key reports to corporate decision makers.

An obsolete search capability forces top managers to make decisions based on outdated and inaccurate information. It can prevent a company from spotting and reacting to a developing trend. A delayed flow of information can contribute directly to financial review problems and compliance difficulties. For customer-oriented organizations, a poor flow of information can contribute to service problems, higher churn, and missed business opportunities.

Beyond the Search Engine

Everyone is familiar with traditional search engines, although many may not understand the underlying technology.

Most Web search engines use automated “crawler” tools to collect every word from virtually every site and page on the Web. That collected information is stored in a cumbersome database table system, and this basic indexing allows the data to be referenced and located when a search is activated. Given the massive and growing volume of information on the Web, this indexing takes place on a vast scale. While the various search engine providers use advanced algorithms and other methods to manage the job, the basic concepts are similar across most search engines.

The ability to find key information is critical for companies today, and in recent years a number of desktop search solutions were created for the corporate environment. Instead of indexing Web-based information, crawler tools are used to search for and organize various types of enterprise content through the use of key words – from internal documents and reports to images, email, corporate intranet sites, and other materials.

Before search-oriented crawlers were introduced to the corporate data environment, companies typically relied on predefined indexes to locate information. The Google-style search allows any word within a document to be automatically indexed, which is why this method is called the “full-text search”. Full-text searching makes every word a searchable value that is easily located when a corporate user types it into a search window.

In today’s dynamic business environment, executives recognize the need to locate and use vital corporate information in a faster and more efficient way than the traditional search engine provides.

Full-text searches bring a number of measurable benefits to the corporate environment. Users can research broad topic areas. Searches typically produce a wide range of results. And because most are familiar with the Google-like approach, very little training is needed to teach employees to use a full-text search system.

But there are also several qualities that limit the usefulness of full-text search systems in the enterprise environment. Because a full-text search returns all values that match the input terms, this approach typically generates a large number of results, most of which may not be relevant or useful to the specific task at hand. For example, searching for “George Washington” in a search engine may bring up institutes and places named after him, links to sites about him, pages of quotations, lists of books about him, and so on. While these are interesting, you may just have wanted to learn what books and articles he wrote. Wouldn’t it be nice if you could just ask for George Washington as an author, and avoid all the irrelevant pages?

Even for local site and intranet searching, having fields of content would make it easier to find the most useful items. Few corporate analysts or decision makers have the time to sort through dozens or perhaps hundreds of query results to locate the precise piece of needed information.

To find and locate highly specific information more quickly and efficiently, a more structured index-based approach should be used yielding a far more focused search result. That focus is achieved by predetermining and defining the values upon which future corporate searches will be based: values such as the name of a customer or an employee, time and location, type of document or other variable. Once an index-driven system is established, corporate users can use those defined value categories, and only those categories, as search criteria.

In today’s dynamic business environment – where financial and competitive pressures, regulatory requirements, and customer demand require executives to make the right decisions more quickly than ever before – astute managers recognize the need to locate and use vital corporate information in a faster and more efficient way than the traditional search engine provides.

Leveraging the “Find Engine”

Results-oriented corporations are not particularly interested in searching for information. They want to find very specific content – quickly and efficiently – and then put that information to work to inform their business decisions, to meet regulatory requirements, and to serve and sell their customer base.

That is why forward-looking business and IT professionals are now looking past the traditional “search engine” model and towards a new and more effective “find engine” approach.

A find engine might best be defined as an extension of and an improvement upon previous generation ECM technologies. It is really the best of both worlds in that it supports and builds upon the established capabilities of the traditional free-form text search, but greatly augments those capabilities with a powerful new suite of automated and index-based capabilities.

The result is a truly new find engine solution that is ideally suited to the dynamic, high-volume data and content management requirements of today’s enterprise.

To gain a better understanding of how find engine technology can benefit the enterprise, let’s examine some of the methods and technologies that drive this new solution.

Multilevel Indexing

The next generation find engines deliver greatly enhanced performance, primarily through the use of multiple levels of indexing. This approach which is far more detailed and comprehensive than traditional corporate indexes acts as a layered table of contents to accelerate and focus find engine results.

Much like a book’s table of contents, a basic index provides a way of logically segmenting and quickly referencing target content in a single document or report. Basic indexing can be applied to a wide range of business documents, including text-based reports, Adobe PDFs, intelligent print data streams (IPDS) such as IBM AFP and Xerox metacode, multi-page image files, as well as Microsoft® Office.

With basic indexing, companies can select and segment important information within a document. By applying index upon index, companies can also drill down into a document to find and deliver very specific bits of data. By granting users very specific authorization to the various segments of a document, organizations can also use local indexing to enforce highly granular document security.

A true find engine:

- Delivers focused results
- Significantly reduces the time spent searching and retrieving information
- Speeds the flow of information
- Improves customer service
- Allows for faster and better-informed decisions

A higher-level index can be used to extend the capabilities of the basic index and to enable tracking and searching any number of index variables across all content within a single repository, and across multiple versions of that content. And finally, there is an enterprise-wide form of indexing which provides a flexible repository for relationship metadata based on business rules established by the organization. Service oriented architecture (SOA) can be used to integrate multiple repositories and index combinations within an enterprise. By issuing queries across metadata repositories, businesses can search vast stores of information and return consolidated results to qualified users.

Other Find Engine Capabilities

The best of this new class of find engine solutions also offer a number of additional query and search capabilities. Many are also offered as a suite of solutions that incorporate traditional full text search capabilities, which are still very useful for broad research and for finding general available information. A reliable find engine needs to go further; be capable of handling both structured data (such as spreadsheets and reports, mainframe output, and application data) and unstructured data (such as Microsoft Word documents and files containing graphics elements).

A fully featured find engine solution should also have the ability to automatically index the internal information in existing documents. This method can be employed, using established parameters to locate, identify, and save index values in existing documents that key triggers during a search of those documents. This type of system allows a true find engine to drill down into the granular details of a document, and produce very specific business information results.

For more complex research applications, a batch query allows a user to build a query containing a number of document segments located in both online and offline data repositories. The system aggregates those document segments into a single work queue entry and then notifies the user when the results are produced. An experienced provider should also craft a solution that presents specific and relevant find engine results through a friendly and intuitive graphical user interface (GUI).

For organizations that possess a large amount of content stored in legacy applications or warehouses, today's most advanced find engines can apply sophisticated automation to index and use that information. An automated legacy archive processor can be used to access, index, and store information (typically most-recent or relevant data) that is needed for current business operations.

In many applications, such as a customer financial statement, a company might need to combine content from segments of various reports into a single printed or electronic output document. A best-in-class indexing solution will facilitate the retrieval of parsed sections from different reports, and from different kinds of reports (such as a PDF, part of an IBM AFP print stream and images from a multi-page TIFF), for repurposing into a single organized PDF for printing or downloading.

Evaluating Find Engine Features

When analyzing potential document and content management solutions, organizations should consider a number of important capabilities.

Scalability - To function in an enterprise-class environment, a search/find engine must support capturing, indexing, storing, and presentation of content at the petabyte level.

Content Management - Content servers can be used to analyze incoming content from structured and unstructured documents and to derive and automatically populate appropriate indexes. The find engine should support multiple local index variables, query constructions, and variable/value combinations. Multilevel indexing can support drill-down capabilities to deliver only those pages or lines that meet all criteria. The best systems also incorporate full-text search capabilities, direct access to archived media, and have the ability to create specific fast-access folders for key business or customer data.

Content Capture - It is important to have a workable search/find solution that can capture all structured and unstructured content including paper documents, email, print streams, images, and all other electronic content easily and effectively.

Security - Any content management solution must incorporate advanced security and audit features, including a single sign-on facility, multiple controlled access levels, robust authentication and authorization protection, and advanced data encryption algorithms. A reliable system will also offer full disaster recovery capabilities.

Benefits of the Find Engine

The find engine approach delivers measurable benefits in an enterprise environment. A true find engine:

- Delivers focused results
- Significantly reduces the time spent searching and retrieving information
- Speeds the flow of information from content stores to customer service representatives, analysts, and corporate decision makers
- Improves service performance and customer satisfaction
- Strengthens regulatory compliance capabilities
- Allows executives to make faster and better-informed decisions

Conclusion

A traditional indexed search typically produces a large number of possible matches, and users must often wade through dozens or hundreds of “results” to locate the intended information. For those that need fast access to very specific information, a traditional search is much like looking for the proverbial “needle in a haystack”.

By contrast, more advanced content management technologies apply multiple indexes and automated analysis to produce only the precise document and item needed. By “organizing the haystack”, this layered indexed approach puts very exact and granular information more quickly and efficiently into the hands of the business user.

It’s the difference between “searching” and “finding”. And it is why a growing number of data-dependent, customer-oriented organizations are migrating to the find engine model.

About Systemware

Systemware, Inc. is a leading enterprise content management (ECM) company. Systemware is able to effectively and efficiently capture, index, store and manage the billions of documents generated by some of the largest organizations in the country. Systemware enables customers across vertical industries—including financial institutions, insurance, healthcare and retail—to make better business decisions, capitalize on information assets, improve customer service, reduce operating costs and ensure compliance. Founded in 1981, Systemware is a privately-held corporation headquartered in Addison, Texas, with branch offices throughout the United States.



systemware

Systemware, Inc.
Dallas Parkway #1100
Addison, TX 75001
1-800-535-8681