

.....

Index Engines Inc.

Pitfalls In Implementing Enterprise Search Solutions

*Enterprise Technology
White Paper*

www.IndexEngines.com

PITFALLS IN IMPLEMENTING ENTERPRISE SEARCH SOLUTIONS

Overview

Today's enterprises face an explosive growth of unstructured information. Corporate employees are burdened with managing ever-larger repositories of files and email—30 times more than they had to manage a decade ago, by some estimates. As a result, they are spending too much time searching for information and not enough time contributing to their organizations' growth and success.

In response to the information glut, businesses are rapidly adopting enterprise search solutions. These solutions are designed to provide quick and easy access to unstructured corporate data. However, there is a fundamental flaw in their approach, which involves using agents that crawl the LAN attempting to access data, often copying it, and then finally indexing it. These enterprise search solutions mimic Internet search engines, which crawl websites in order to find new content. But what works on the Internet does not succeed within the confines of the enterprise. The agent-based approach adds complexity and inefficiency to an already overburdened network.

...the correct approach is to integrate the indexing process tightly and transparently into the existing storage fabric.

A new approach is needed—an approach that solves the problem from a different angle and delivers the benefits of existing enterprise search technologies without any of the pain. Index Engines believes the correct approach is to integrate the indexing process tightly and transparently into the existing storage fabric. When indexing is a transparent part of existing storage processes, it will be integrated into existing storage security mechanisms. This approach, when properly executed, adds no additional work, performance impact, security vulnerabilities, or network resource consumption, but instead provides a powerful enterprise search solution to even the most casual user in the enterprise.

This white paper focuses on issues to be aware of when implementing an enterprise search strategy. The companion white paper, Rethinking Enterprise Search, presents Index Engines' innovative new solution to the challenge of enterprise search. It is available upon request at the Index Engines website, www.indexengines.com.

1.1 The Information Explosion

In the days when few employees had personal computers and even fewer used email, it wasn't difficult to find information in electronic files. Today's PC users have a much harder time. According to technology market research firm The Radicati Group, the average knowledge worker sends 34 emails and receives 99 emails daily, for a total of 14.7 megabytes of email per day, or 53 percent more than last year. This adds up to an astounding 29,000 emails per year, more than 3 gigabytes of data for each user.

Corporate data now increases by 50 percent to 70 percent every year.

User files such as documents and spreadsheets have also proliferated. According to the IT research firm META Group, the average corporate storage per employee has grown from less than 100 megabytes in 1993 to more than 3 gigabytes in 2003—about 18,000 files, assuming an average file size of 175 kilobytes.

This information glut may even be accelerating. Consulting firm Horison Information Strategies estimates that corporate data now increases by 50 percent to 70 percent every year. The typical employee managing 18,000 documents today may have 46,000 documents to manage in two years. Knowledge workers will spend more and more of their time navigating a complex maze of documents and email messages, searching for the information they need to do their jobs. The threat of work grinding to a halt as employees wade through a sea of information is leading many firms to put a high priority on implementing enterprise search.

1.2 Why Tame the Beast?

Getting a handle on corporate information—allowing users to find what they need quickly and easily—is a difficult task. User behavior, complex networks, disparate databases, and diverse file types make the challenge a complex one for even the most seasoned IT professional. But the benefits of taming the information beast are enormous.

Research firm IDC found that knowledge workers typically spent between 15 percent and 30 percent of their time looking for information in their files. What's more, the employees reported that they were unsuccessful at least 50 percent of the time, and had to redo the work they had lost. If we estimate, conservatively, that each employee spends an hour per day looking for information and half an hour per day re-creating lost information, the losses to a company with 1,000 workers could easily amount to \$15 million per year:

...each employee spends an hour per day looking for information and half an hour per day re-creating lost information

Average annual employee cost (fully loaded):	\$80,000
If a typical employee spends one hour each day looking for old files and email:	\$10,000
And a half hour per day reworking information they cannot find:	\$5,000
Total cost per employee:	\$15,000
Total companywide annual expenditure for searching and reworking enterprise data (1,000 employees)	\$15,000,000

Table: Annual Employee Expense

But this estimate includes only the cost of the time wasted by employees. It doesn't include the value of the work the employees might have done in that wasted time. Corporations hire employees in the expectation that they will provide value in excess of their salaries. If we assume, conservatively, that a typical employee should be providing twice his or her salary in value to the corporation, \$15,000 in lost time actually represents \$30,000 in lost value. Therefore, for a company with revenue per employee of \$250,000, wasted time per employee can amount to more than 10 percent of total company revenue. The return on investment (ROI) of minimizing this lost time and enabling employees to generate more value for the company can be considerable.

2.1 Where to Begin?

...look to the SAN to manage the process of finding and aggregating... then tap into this process to accomplish enterprise search.

To solve the enterprise search challenge, it seems natural to start where the information resides: in the data network. After all, the LAN knows where files are located, when they are updated, and who can access them. Because the LAN controls the data flow of the organization, it seems like the logical place to begin.

However, the LAN is a complex and overburdened environment. Most organizations find it an enormous challenge to keep the network flowing and to allow access to information at the speed of light. Why add more complexity by layering on enterprise search, which is resource-intensive and complex, and which requires continual network crawling in order to catalog information?

The storage area network, or SAN, may offer a better opportunity for solving the enterprise search challenge. We can look to the SAN to manage the process of

finding and aggregating files and email, and then tap into this process to accomplish enterprise search.

In the remainder of this paper, we discuss some of the challenges in implementing a LAN-based solution. Following this review, we will look at a fresh architectural approach to tackling the enterprise search challenge.

2.2 Users Aren't the Problem

Requiring them to learn complex new systems, and spend hours in training, is not the most productive use of their time.

One way to approach the enterprise search challenge is to start with the users. After all, they're the ones who have created the problem by producing all this information! We could make information easier to access by changing user behavior. Defining document policies and procedures, or requiring users to load documents into predefined repositories, would be an ideal solution—as long as the users abide by these new rules and correctly store all their information, including historical information.

This approach seems to work for very well-defined environments where information is predictable and manageable. However, the majority of corporate information does not reside in structured repositories, nor is it managed in well-defined file structures. In fact, according to many industry analysts, unstructured data accounts for over 80 percent of a firm's total information assets.

Even the most comprehensive and feature-rich information repository will be useless if users do not adopt it and support it. Users already feel overwhelmed by technology. Requiring them to learn complex new systems, and spend hours in training, is not the most productive use of their time. Therefore, Rule #1 for selecting an enterprise search strategy is: Make it easy for users. Do not require them to change their behavior.

2.3 Integrate and Unite!

...enterprise search solutions require significant customization before they become productive

Most enterprise applications today require tight integration into an already complex environment. Knowledge management, customer relationship management, and sales force automation systems—to name just a few examples—all find data, capture it, and link to it throughout the enterprise. While this is a commendable approach, it requires significant implementation and customization efforts on the part of the IT organization. Many of these projects can take six months or longer before they are launched and become productive. This places a significant burden on an already resource-constrained IT group, and may require consulting fees equal to or even double the cost of the software itself.

Enterprise search solutions are heading in the same direction as other business applications. Many of today's enterprise search solutions require significant customization before they become productive. Some actually require that all data be in a specific format, or that it be web-enabled before it can be found and indexed. Rule #2 for your enterprise search strategy is: Minimize

implementation and customization requirements—or better yet, eliminate them. Let your enterprise search solution, not your IT department, do the heavy lifting.

2.4 Leap, Don't Crawl

...crawling the network to find files and email, not once a day, but continually throughout the day, is a flawed foundation for an enterprise search strategy.

As Internet content exploded, search engines began aggressive efforts to crawl this content so they could provide users with the tools to find relevant information quickly and easily. Over many years, they solved this problem; today, finding anything you need on the Internet is painless. Enterprise search solutions are now using the same paradigm. However, an enterprise environment is not analogous to the World Wide Web, and attempting to retrofit web search into enterprise search is a mistake.

It's true that enterprise search solutions need to find data in order to index it. But crawling the network to find files and email, not once a day, but continually throughout the day, is a flawed foundation for an enterprise search strategy. Typical infrastructures have been in place for many years and are feeling the burden of the massive information explosion. Email, documents, Internet access, and now instant messaging are all taking a toll on the network. Network crawling adds tremendous traffic to the network and strains an already fragile environment, impeding data from moving freely. Rule #3 of your enterprise search strategy is: Don't add any LAN traffic or steal CPU resources.

2.5. Leave Me Alone

Many enterprise applications require significant care and feeding.

Your technical experts spend most of their time administering existing systems. Maintaining the current environment so that all systems are up and running is a full-time job. Many enterprise applications require significant care and feeding. Also, every new application brought in-house imposes a learning curve on IT professionals.

Administrative tasks aren't the only time-wasters for the IT staff. Training users and installing desktop software also require considerable resources—not just when the system is first implemented, but whenever it is updated or maintained. As new employees join the company and as new software versions are released, this distributed environment becomes a nightmare for systems engineers. As a result, many firms are often several versions behind in their major systems upgrades.

Rule #4 of your enterprise search strategy should be: Don't burden IT staff or users with ongoing maintenance, installation, or training. From day one, the system should run independently, with little or no intervention from IT. User training should take three seconds or less.

2.6 Grow with Me

...a solution that has been customized ... for a specific departmental environment may not work outside that domain.

Companies often begin by implementing enterprise search in a single department, and then find that expanding to a corporatewide system is complex and time-consuming. An enterprise search solution that has been customized and implemented for a specific departmental environment may not work outside that domain, and may require software or customization at additional expense.

Rule #5 in selecting an enterprise search solution is: Make sure the system is scalable. Whatever work it takes to implement the system in one department will have to be repeated when the implementation is expanded to other departments. Also, you should understand what software and hardware are required to support additional users.

3.1 Storage Networks versus Data Networks

For all of these reasons, introducing a sophisticated process such as enterprise search into the heavily trafficked, business-critical, and overburdened environment that is the data network may not be the most logical approach. A better approach to enterprise search would be to index data automatically via the storage side of the business infrastructure, invisibly to the user. This approach eliminates the need for agents to crawl through the LAN, seeking data. It requires little or no attention from end users or IT professionals, and does not affect performance or security on the network.

3.2 Conclusion

Implementing an enterprise search solution for your organization can yield tremendous return on investment and make users significantly more productive. However, choosing the wrong platform for this solution could cost you valuable time and resources, and be a source of frustration.

When researching solutions for enterprise search, remember these five rules:

1. **Keep it simple for your users.** *Since user adoption is critical, do not give users a complex solution that requires them to change their behavior. Provide them with a simple solution that gives them the answers they need when they need them.*
2. **Keep it simple for your IT group.** *Do not go down a path that requires long-term planning, significant customization, and complex implementation. You can successfully implement a plug-and-play enterprise search strategy that does not require any IT resources.*
3. **Do not burden the already overtaxed data network.** *Crawling the network to seek out data is not a solid foundation for an enterprise search strategy.*
4. **Implement a solution that does not require maintenance or monitoring during day-to-day operations.** *User support, maintenance, and training should be nonexistent.*
5. **The solution should be scalable with no additional hardware, software, customization, or training.** *Choose a system that will easily and cost-effectively grow with the organization.*

Enterprise search is an evolving area. Existing technologies have been in place for years, or have evolved from existing enterprise applications. Burdened with this legacy, they are locked into a fundamentally wrong approach. By taking a step back and considering a logical, efficient, and a cost-effective approach, you can make a radical departure from today's solutions.

The Index Engines approach to enterprise search is presented in the “Rethinking Enterprise Search” white paper. This white paper can be requested on our website: www.indexengines.com.