

**Information Management** software



## Automating daily document decisions with IBM Classification Module.

*Best practices for deploying automatic classification in your organization and optimizing results*

Filing documents is rarely the best use of a knowledge worker's time. It's costly, tedious and burdensome. Who has time to stop productive work for every document created or every e-mail received or sent? Who can take the time to determine where documents should be stored—or when they should be trashed?

But interrupted work and lost productivity are only the beginning. Manual filing can lead to consequences that are even more costly to the business. Slow access to information can delay business decisions, from daily details to major initiatives. And inconsistent execution of disposition policies or an inability to locate and produce content in a timely manner once it's requested can lead to legal or regulatory challenges.

### Using technology to relieve workers of the classification burden

As more companies discover that manual filing techniques developed in a paper-based world are not viable in electronic environments, an increasing number are turning to technology for help.

Software solutions such as IBM Classification Module organize documents automatically using a combination of natural-language processing and statistical analysis. IBM Classification Module examines and analyzes the full text of electronic documents and e-mails to generate suggested categorizations. Not only does the system train itself with actual examples from your organization, it can learn from user feedback and incorporate that feedback into the system's understanding in realtime. The result is a system that can consistently determine the proper categorization for content based on the full context of the document.

But the software can't do its job alone. Deploying an automated classification system requires the company to implement it using best practices and business processes if the solution is to achieve its intended result.



## What's driving automatic classification?

Multiple factors are driving the need for automatic classification. Part of the challenge is the difficulty and cost of enforcing manual policies and procedures in the face of growing content volumes. For example, employees are often wary of following manual records-classification procedures. Many users consider the additional processes burdensome and refuse to participate. And companies sometimes find that the requirements of their compliance initiatives limit participation by workers.

Though humans are still the most accurate decision makers when motivated to do the job, factors such as these make them the least-reliable solution for widespread content classification in your organization. When potentially low levels of participation are considered, humans provide the least reliable classification results.

Compliance efforts only add to the challenge because employees have to answer more questions about everything they produce. Is this a record? Does it have business value? What is the retention policy for this type of file? How should I file this? Moreover, the monetary cost of daily participation from workers is high.

A knowledge worker earning US\$75,000 per year who takes 10 seconds to manually classify information can incur as much as 17 cents per document<sup>1</sup>—an expense that adds up fast when you're talking about thousands or millions of files.

And with unstructured content being generated in the enterprise at higher rates every year, today's mounting productivity costs can only grow. With every e-mail, instant message and piece of business content representing a potential classification decision, even modestly sized organizations can calculate significant increases in cost by forcing end users to make compliance decisions on content.

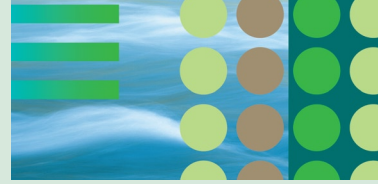
By contrast, an automated classification system such as IBM Classification Module can enhance and simplify content-centric solutions by helping ensure that documents are filed in appropriate and predictable locations—without burdening your organization's workers. An automated filing system also can be a key building block in creating decision-making processes that are themselves automated—where the availability of information can trigger steps in a business process without the need for human intervention.

## Overcoming objections to a software solution

Despite compelling reasons for adopting automatic classification, many companies are reluctant to adopt the solution. Many fear consequences from an inaccurate suggestion.

This worry affects all potential uses—whether in day-to-day access and decision making, records management for regulatory compliance, or records retention to guard against legal action. Many companies fear that they may have to defend their technologies in court or before a government agency—and they are unsure whether automatic classification can hold up to scrutiny.

The answer is that technology and the law can, in fact, be compatible. Courts typically are not concerned with technology alone. Rather, they are interested in how companies use the technology, the effectiveness of the technology and—most important—the results that it produces. A key step for users, therefore, is to have good reasons for applying technology and to select a solid and proven technology—such as IBM Classification Module—that helps produce consistent, high-quality and accurate results.



## **Best practices for getting started**

A company that deploys a content classification solution can directly affect how well the solution functions and how well it meets compliance and legal needs. Here are some of the best practices you should follow as you begin working with automatic classification.

### *Optimizing investments*

The laws of civil procedure do not require companies to file or organize their content in a certain way, but your approach can have a significant impact on long-term costs and success. Up-front “training” of the system, using a rich set of example documents for specified categories, is essential to successful use of learning-based classification approaches. Most organizations find that the extra investment in gathering and applying training content early on can pay off significantly over time with high accuracy—especially compared to the cost of manual oversight.

### *Balancing business needs versus risks*

To balance the tradeoff between accuracy and automation, it’s important to perform a thorough risk analysis. From a high level, companies have a choice to implement highly automated processes, acknowledging a tolerance for

potential mistakes. Or they can limit the amount of automation to specific areas and more closely monitor processes. The risk analysis will need to incorporate the reality that humans are unlikely to consistently participate across the enterprise in your program. Where you draw the line along the automation continuum should depend on your tolerance for risk given your specific business scenario and the associated cost of alternative methods.

### *Keeping it simple*

With many compliance initiatives, it is neither practical nor smart to start “big.” IBM clients and Forrester<sup>2</sup> point out that it’s best to start with a high-level approach to creating your automated records management file plan or taxonomy. Once your initiative has gained control over your content with a coarse organization of content, you can begin to add granularity over time—in an iterative fashion that enables you to more easily phase in your solution and limit risk.

### *Establishing processes*

With any automated solution, it’s a good idea to conduct a pilot with a small group of users to better ensure the effectiveness of new processes. Classification is no different. Because an element of uncertainty is introduced—classification

responses come with a variety of confidence levels—a pilot is a recommended best practice. Although it’s impractical to review a large percentage of decisions once a solution is deployed to the enterprise, looking at the classifications for a select pilot group is manageable—and provides the opportunity you need to fine-tune the training of the system. What’s more, over time, batching together adjustments and changes and piloting them with select users before a general rollout can help you find the right balance when it comes to choosing your automation level.

## **Best practices for ongoing operations**

Content classification—as is the case with any automated solution—requires ongoing oversight to establish that your classification results are consistently good. Here are some best practices you should follow as you continue using your classification solution.

### *Routine consistency*

One key to effective management of electronic records and documents is consistency. And to ensure consistency, a critical practice is to make your operations as routine as possible.



If you are consistent with the way you apply technologies such as classification, your document filing is more likely to be accurate and to constitute a form of complete disclosure that is necessary to make your practices less vulnerable to challenge. The goal is to achieve what is known as “routine good faith operations”—a consistent and honest attempt to file and protect what is necessary. Such practices may grant you protection from legal action even if documents are inadvertently destroyed or not retained. And automated classification, with its repeatable algorithms, is consistent—especially when applied in a routine, scheduled manner and especially when compared to the unpredictable behavior of the average knowledge worker.

### Audit and review

To help ensure the continued effectiveness of your content classification solution, you need to conduct regular audits of its operation. Keep an eye on the accuracy of results by monitoring not only the documents saved but also the documents rejected. Randomly sample results to ensure that they meet your business standards and the standards of your compliance initiative. This auditing process also gives IBM Classification Module an opportunity to learn in realtime—the feedback provided as part of an auditing process can become the basis for a practice of continuously improving the system’s understanding.

Beyond executing ongoing auditing of your results, you also should document your auditing processes themselves, the results of audits and any actions you have taken to remedy issues or improve accuracy. By documenting your processes, you can demonstrate that your automated classification solution is being properly overseen, and provides accurate results as part of routine and effective operations.

### For more information

For more information about IBM Classification Module, please contact your IBM sales representative or IBM Business Partner, or visit:

[ibm.com/software/ecm/classification](http://ibm.com/software/ecm/classification)

© Copyright IBM Corporation 2008

IBM Corporation  
Software Group  
3565 Harbor Boulevard  
Costa Mesa, CA 92626-1420  
U.S.A.

Produced in the United States of America  
August 2008  
All Rights Reserved

IBM, the IBM logo, and [ibm.com](http://ibm.com) are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at “Copyright and trademark information” at [ibm.com/legal/copytrade.shtml](http://ibm.com/legal/copytrade.shtml)

Other company, product, or service names may be trademarks or service marks of others.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates.

The information contained in this publication is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this publication, it is provided “as is” without warranty of any kind, express or implied. In addition, this information is based on IBM’s current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this publication or any other materials. Nothing contained in this publication is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software.

Each IBM customer is responsible for ensuring its own compliance with legal requirements. It is the customer’s sole responsibility to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer’s business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer is in compliance with any law.

1 Cohasset Associates, Inc., *IBM FileNet Records Manager: Cost Effective Electronic Records Management, An Assessment*, November 2007.

2 Murphy, Barry, *eDiscovery Best Practices for Information and Knowledge Management Professionals*, Forrester Research, Inc., September 24, 2007.

