Brochure

Identify and Automatically Declare Valuable Information for Better Governance

opentext™

Identify and Automatically Declare Valuable Information for Better Governance

Many organizations find that even with enterprise content management systems in place a percentage of their information doesn't make it into the system and sits ungoverned in file shares, SharePoint sites or other repositories. You need to know where all your valuable business information resides and be confident it is secure and managed appropriately.

It's a Challenge to Manage All Information According to Policy

The reality today is that you cannot guarantee all of your valuable information is being managed within the appropriate enterprise systems. Information is created and captured from a variety of sources, in a variety of formats by most staff within your organization. Many "information workers" today do not have the training, skills, or interest to manage business information in accordance with corporate mandates. Your staff are trying to get their job done as quickly and as simply as possible, which often means valuable business information may be stored in file shares or SharePoint sites outside your governance controls.

OpenText™ information governance and file analysis solutions can help you address this unstructured data management challenge by connecting and managing the information in your business systems better, giving you greater control over this valuable asset.

Auto-Categorization and Declaration Can Help You Govern Information Better

Organizations collect a vast array of electronic content including documents, messages, and records (transactions, staff files, and contracts, etc.). Much of this information is valuable to the business and subject to regulatory requirements. For example, personnel files must be kept secure and maintained for relatively long periods of time or at least while the individual is employed by the organization. Customer credit card details, however, captured as part of a transaction may need to be deleted from system records within a shorter period to comply with privacy legislation. In many cases, regulations drive the need to secure and appropriately manage this business information, but what about intellectual property such as product designs and roadmaps, future expansion, and acquisition plans? This intellectual property may not be covered by global regulations but it is equally important to manage it appropriately and ensure it is secure and easy to find.

Categorization is critical to the application of appropriate policy which controls how information is governed. Traditional collaboration and content management systems rely on users to manually categorize and tag electronic content on an individual basis, which is proving to be increasingly time-consuming and error-prone. The ability to identify and auto-declare this valuable information is like having a safety net to catch anything that slips through the cracks, and being part of your information governance strategy makes it easier to manage your content to improve your compliance state and reduce risk and cost.

Let ControlPoint Help You Identify, Categorize, and Declare Information

With the large volume of content under management today, you need to reduce your reliance on staff putting that content into appropriate folders or managed repositories. You also want to be confident that you can find any valuable information in siloed repositories quickly and efficiently. OpenText™ ControlPoint can help you identify, analyze, categorize, and declare diverse types of content stored in enterprise repositories by leveraging the OpenText™ IDOL connector framework.

The Categorization Pipeline

A simple way to think of the categorization and declaration process is in terms of IDOL's categorization pipeline, illustrated in Figure 1. Basically IDOL reads the documents and gathers the words, it then sorts the words out, groups them together and matches them against known templates. Once matched an associated action is performed.

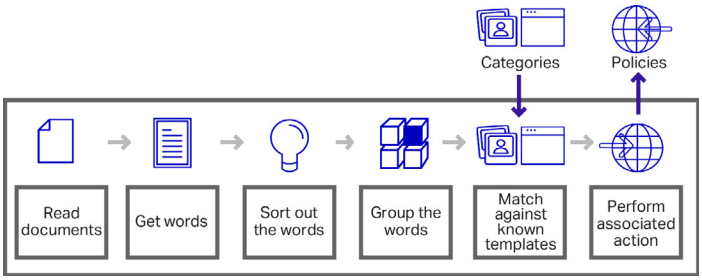

Figure 1. Simplified IDOL categorization pipeline

When this is translated to ControlPoint's auto-declaration capability, the "known templates" are defined categories and the "associated action" to be performed on the content is the application of appropriate policy that has been defined in ControlPoint. In this manner, ControlPoint is able to determine if specific content needs to be managed and applies policy to make this happen.

So how are categories defined in ControlPoint and how does it know what the content should look like?

Streamline Category Definition through Training

To improve the efficiency and accuracy of categories and the application of policy to content, ControlPoint categories can be trained on a range of file analysis parameters. These include training text extracted from a sample document, field text taken from the document or its location, or training documents that are a set of sample documents chosen on the basis of their content. By aligning the categories with meaningful concepts and real business context, ControlPoint removes much of the burden of having users manually create or map these categories. An added advantage for those organizations who have Micro Focus Records Manager is that ControlPoint categories can be trained on the business classification scheme and selected records.

Categories can be refined and tested to determine the impact they are likely to have on enterprise documents by adjusting the weighting of a term, the selection of threshold values, or by adding field text.

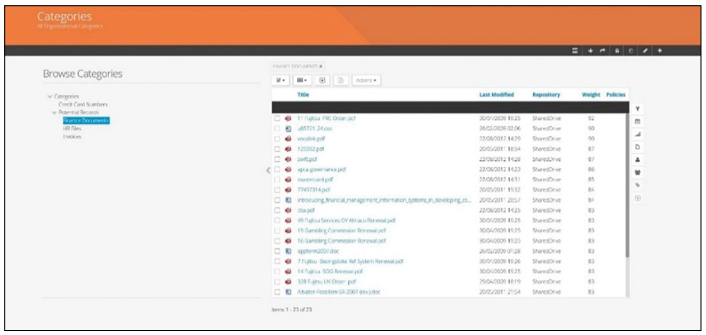

Figure 2. Browsing categories in ControlPoint

These sophisticated file analysis activities can be done individually or in combination, and when you are happy with the resulting category you can publish it and make it available for use in automatic policy execution against electronic content managed by ControlPoint.

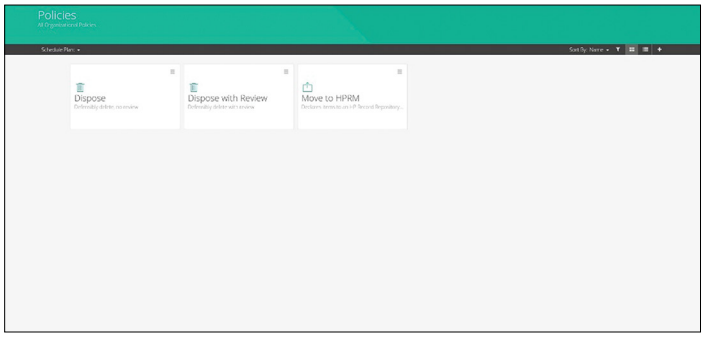

Figure 3. Policies page in ControlPoint

Identifying Sensitive Information to Improve Security and Reduce Risk

One of the greatest risks facing organizations today is the unwitting retention of sensitive and personal information within system repositories. Personal Credit Information (PCI), Personally Identifiable Information (PII), and Personal Health Information (PHI) need to be managed in accordance with privacy legislation to protect the individual. There have been numerous publicized cases of large corporations being hacked and personal information or credit card details (that should have been deleted) stolen and used for blackmail, extortion, or threats. Not only is the loss or leakage of this sensitive information a serious threat for the individual it pertains to, but there are wide-ranging consequences for

Connect with Us
OpenText CEO Mark Barrenechea's blog

the organization holding the information. Consequences include fines or sanctions, reputational damage for the organization (and its executives), and the potential collapse of the business.

This sensitive and personal information should not be held in file shares, SharePoint sites, mail servers, or unsecure system repositories. However, it is surprisingly common to find it there.

ControlPoint and its file analysis algorithms help you detect potentially sensitive information during repository scanning by leveraging IDOL categories and eduction grammars. Eduction provides predefined grammars that identify and extract information in certain formats such as credit card numbers or addresses. Items of interest can be displayed on the summary dashboard for easy reference. The identification of "items of interest" is based on category matching and may be defined as documents that contain personally identifiable information such as social security numbers or email addresses.

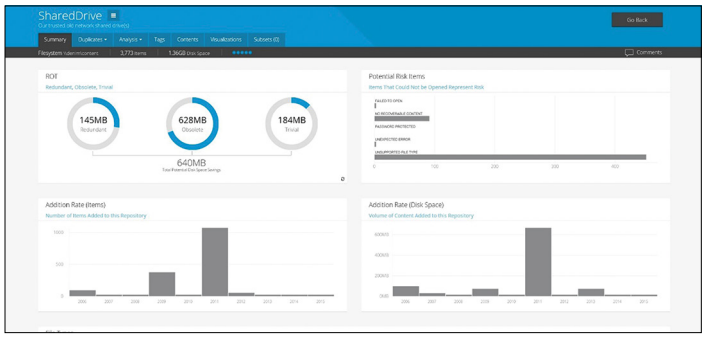

Figure 4. ControlPoint Statistical summary dashboard

Apply Policy to Take Action and Manage Electronic Content

Once ControlPoint has identified and categorized your content, it can automatically apply the appropriate policy to facilitate the ongoing management of this content. A ControlPoint policy defines the rules and actions to be performed on managed content and these actions include: declare, declare in place, dispose, hold, release hold, no action, secure, tag item, and update. For example, a ControlPoint policy may be defined to secure correspondence relating to supplier contracts in a SharePoint site and then delete it five years after the date of creation, or a policy could declare business records relating to health and safety into a Records Manager repository. The ability to assign policy to content automatically based on its categorization removes this often time-consuming and error-prone process from staff helping you improve productivity and compliance with regulatory requirements.

Close the Loop on Information Governance by Auto-Declaring Records

In addition to standard policy actions that can be applied to electronic content in different repositories, ControlPoint has specialized file analysis capabilities that enable business records to be identified automatically within an organization's content and declared intelligently into a Records Manager repository. Business record identification and filing continue after the initial indexing via Records Manager's auto-classification capability, so new content entering the organization, which matches the record identification rules, is captured automatically as a business record.

The ability to identify and categorize your enterprise content from the redundant, obsolete, and trivial (ROT) data taking up storage space through to the sensitive information and business records that are vital to business operations as part of an end-to-end information governance solution delivers significant benefits to your organization and staff. Combining proactive content management with the ability to retrospectively categorize and declare unmanaged content helps you close the information governance loop to benefit from:

- Improved records capture and regulatory compliance
- Improved information security and risk mitigation
- Greater ability to identify sensitive content and records sitting amongst ungoverned content
- Reducing the burden on staff to manually process and tag content
- Reduced cost to store and manage content

Learn more at
www.microfocus.com/infogov
www.microfocus.com/opentext

opentext™