

CONTENT MANAGEMENT

SOLUTION BRIEF

Context

Challenges and Insights

ABBYY Smart Classifier

The Challenge of Big Content

We live in the age of Big Data, with unprecedented volume, velocity and variety of data. Business process and content management used to give us what we needed to manage our enterprise content, even if the results weren't fully optimized. As Big Data continues to grow exponentially, the requirements of business process and content management have been transformed inexorably into supporting this new world of Big Content.

Information overload is real, and it can have a deadly effect on workplace productivity. Most - if not all enterprise workers - are "knowledge professionals." They work with information continuously, connect information to their own knowledge, and constantly make decisions. In our age of Big Content knowledge professionals no longer face an issue of too little information; instead their challenge is needing to assimilate too much - far, far too much - information. Automatic text classification organizes and prioritizes information so that knowledge professionals have the information they need - exactly.

DATA MIGRATION

Challenges and Insights



Nobody Organizes Big Content – It’s Just Too Big

Information management prior to the age of Big Content was often handled manually: manual-creation, manual classification, and manual organization. It didn’t always work perfectly, but it worked. But the volume, velocity and variety of Big Content changed all of that. In a Big Content universe our fundamental business need to organize, prioritize and efficiently access information means that time consuming, inaccurate and inconsistent manual classification is simply unworkable. **Automatic text classification enables knowledge professionals to efficiently locate information critical to their work.**

How to Re-empower Search

Out-of-the-box keyword-based search systems provide knowledge professionals disorganized, seemingly random results. Knowledge professionals need the right information – exactly, at the right time and in the right context. Knowledge professionals expect to get information from your enterprise systems in the same way that they get it from the Internet: quickly, easily and prioritized with what they need first. But to deliver that kind of simplicity requires complicated metadata on the backend to structure otherwise unstructured content. **By recognizing the meaning within Big Content automatic text and semantic classification can provide the missing metadata to organize Big Content so knowledge professionals can search and retrieve critical content fast.**

Managing Content for Knowledge Professionals

Your organization’s knowledge professionals are subject matter experts; that’s what lets them analyze information based on their knowledge and experience to make the right decisions. To expedite their work they need to be able to put their hands on pertinent documents within Big Content immediately. so as to organize the information around their expertise. **High-performance content classification organizes multi-faceted search results based on document content so that knowledge professionals can focus in and drill down on the results that they know they need. And by efficiently organizing Big Content knowledge professionals can easily navigate a taxonomy of categories to locate relevant documents.**

DATA MIGRATION

Challenges and Insights



Identify Policy Violations And Uncover Hidden Risks

Big Content that is out of control (or out of sight) can bring big risks. Automatically classifying content can help you identify patterns of policy violations from data that is being misused or misplaced. It can also help uncover risks across data that is being stored but no longer accessed or needed by anyone. **You can then create a baseline for future data discovery and define a set of key risk indicators (KRI) and key performance indicators (KPI) to find data silos that are outside of the standard business activity.**

Keep Big Content Under Control

Metadata-based disposition rules, whether using dates, custodians, or file type are blunt instruments, and unlikely to satisfy Legal or Compliance scrutiny. **Automatic content classification combining both high performance text-based classification and powerful semantic based classification enables you to identify data that should be discarded or archived at a targeted, granular level.** You can efficiently de-duplicate, as well as detect and dispose of “ROT” – Redundant, Obsolete and Trivial data to help maintain the health of your Big Content repositories. If you are challenged in a legal or regulatory proceeding you can demonstrate that your disposition was reasonable, using policies based upon a clear understanding of the data.

Automatic Text Classification

To move beyond manual classification and data organization requires classification technology that can automatically categorize and prioritize content, Traditional classification approaches use document metadata, for instance the MS-Office Property sheets, or headers in email. But using metadata to organize Big Content is too crude to power granular, efficient search and discovery. We need technology that can dig deeper and categorize documents based on their meaning and/or key characteristics.

Text and semantic-based classification technology that automatically understands the meaning, key facts and events from a document can deliver granular categorization that is required to get Big Content repositories under control, enabling knowledge professionals to find and locate documents that are critical to their decision-making activities.

ABBYY SMART CLASSIFIER

Intuitive and easy-to-use

Based on a powerful engine that provides both text-based and semantic-based classification, ABBYY Smart Classifier is designed to make classification a snap - for anyone - Information Managers, Business Analysts, and even technical wizards!

The design of ABBYY Smart Classifier marries high-performance, high-accuracy classification algorithms to features that radically simplify the classification workflow process to overcome the barriers-to-deployment confronting too many enterprises today.

What makes ABBYY Smart Classifier different?

Information Management professionals need to be able to quickly stand-up taxonomies and classification models for Big Content to find and deliver the right documents - exactly. ABBYY Smart Classifier radically simplifies the classification workflow process to accelerate the time-to-deployment of high performance, high accuracy text classification.

Model Editor

Smart Classifier's Model Editor is an intuitive and easy-to-use interface that guides line-of-business managers through the taxonomy and classification model development process. By providing direct feedback and metrics, it enables users to quickly develop, stand up and tune taxonomies and classification models.

Automatic Optimization

The accuracy of automatic test classification is dependent on a large set of statistical and semantic parameters. Tuning these manually can be a daunting task. ABBYY Smart Classifier does this dirty work for you - automatically making the decisions to optimize the algorithms and deliver consistently high classification quality.

Text and Semantic-based Classification

Based on Compreno, ABBYY's innovative natural language processing technology, Smart Classifier can classify documents based on the linguistically derived meaning of words and sentences, including relationships, facts and events.

Small Training Sets

Classification models are "trained" using sample documents. In the past this required thousands of documents to train a category. Smart Classifier reduces that by a factor of 10 or more, minimizing the burden on IT and knowledge professionals.

Simple APIs

Using a simple REST API, integrating Smart Classifier into workflow, archives, records management systems, email management, data migration or compliance solutions is a snap - efficient, fast and easy.

ABBYY SMART CLASSIFIER

Intuitive and easy-to-use

Take control

Being able to quickly locate and find pertinent documents for decision-making activities within Big Content repositories depends on the accurate assessment and understanding of each and every document. Automatic text classification is your primary weapon to organize and prioritize Big Content to support your knowledge professionals as they make critical business decisions. With ABBYY Smart Classifier you can drive the process yourself from start to finish - from defining your taxonomy categories to developing and tuning your training models - breaking free of IT resource constraints and long development cycles.

Classify with confidence

ABBYY Smart Classifier provides both high performance text and semantic-based classifiers. The Model Editor enables you to tune each category definition to a high level of accuracy. Randomly constructed Control Set documents are used to identify true/false positives and negatives, and precision, recall and f-measure metrics enable you to manage the behavior of your model. These inputs, all of which are presented and can be managed in the intuitive Model Editor, enables you to craft classification models best suited to your governance requirements.

Use project based workflows to manage the classification process from model creation to deployment.

Analyze and tune models using in-depth category metrics, including f-measure, precision/recall, true and false positives/negatives and more.

View features of documents in training or control sets and tune models based on key characteristics.

ABBYY SMART CLASSIFIER

Intuitive and easy-to-use

The easy-to-use Model Editor enables you to manage the classification process from model creation to deployment. View features of documents in training or control sets and tune models using in-depth category metrics, including f-measure, precision/recall, true and false positives/negatives and more.

Training Set

Sample documents used to train a mathematical model to classify into categories

Model

Unstructured data classification
with ABBYY Compreno
model test result

Control Set

Analyze and Tune

F-Measure / Precision / Recall
Features
Inclusiveness
True Positive / False Positive
True Negative / False Negative
confidence and probability

Deploy to
Production

Key Features and Benefits

ABBYY Smart Classifier is explicitly designed to simplify classification workflow so that you can take advantage of high-performance, highly accurate text classification to mitigate risk and implement effective governance policies

Workflow Productivity

- A single integrated environment to manage all aspects of classification
- Multiple taxonomies and models to enable iterative development and multiple business processes
- Easy-to-use Model Editor to guide users through classification workflow

Languages and Document Formats

- High-performance text-based classification in 39 languages
- Enhanced semantic-based classification in English and Russian
- Support for all major business text document and image formats

Model Optimization and Performance

- Taxonomies, training sets and models are created and managed with the intuitive Model Editor
- Inclusiveness of classification models is designated at model creation and tuned during development
- Classification models can be tuned using performance metrics such as f-measure, precision/recall, true/false positives, and more
- Training sets can be optimized based on granular document features and characteristics
- Using Control Sets ensures that the performance of classification models can be measured and tuned independent of training bias



**ABBYY International
Headquarters**

Otradnaya str. 2b/6 127273
Moscow, Russia

Tel: +7 495 783 3700
Fax: +7 495 783 2663

Email: office@abby.com

**ABBYY European
Headquarters**

Elsenheimerstrasse 49, 80687
Munich, Germany

Tel: +49 89 69 33 33 0
Fax: +49 89 69 33 33 300

Email: sales_eu@abby.com

ABBYY Australia

Citigroup Building, level 13,
2 Park Street
Sydney, NSW, 2000
Australia

Tel: +61 (02) 9004 7401

Email: sales@abby.com.au

**ABBYY North American
Headquarters**

880 North McCarthy Blvd.,
Suite #220
Milpitas, California 95035
USA

Tel: +1 408 457 9777
Fax: +1 510 226 6069

Email: sales@abbyusa.com

**ABBYY 3A
(Asia, Africa, South America)**

Otradnaya str. 2b/6 127273
Moscow, Russia

Tel: +7 495 783 3700
Fax: +7 495 783 2663

Email: sales_3A@abby.com