MICRO FOCUS

Data Sheet

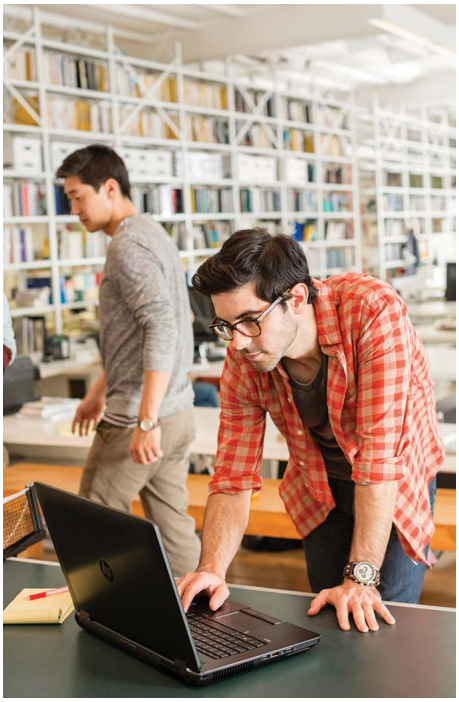
Information Management and Governance

KeyView

Empower your customers to get more out of their data by providing accurate file format detection, content decryption, text extraction, subfile processing, non-native rendering, and structured export, with support for over 1,500 formats across all major client and server-side platforms.

Product Highlights

Today's businesses need maximum visibility into their data to stay competitive. However, they struggle to manage the wide and ever increasing variety of document formats arising from digital transformation, from those found in legacy repositories (which current software cannot read), to new applications, software updates, and nascent cloud-specific formats. These challenges drive the need to convert information into a consistent format to make effective use of the information within. Micro Focus KeyView is an embeddable content extraction and file transformation solution for leading edge software developers and service providers, which not only adds business value, but also significantly reduces business risks associated with managing large volumes of information.



The Challenges of Information Access

It can be difficult, expensive, and sometimes even impossible to ensure that you have the appropriate applications or plug-ins required to handle a multitude of document formats. With such a wide variety of file formats and data types, having a technology that simply connects to repositories through connectors is not enough. Once connected, it must also have the proper file filtering technology to allow access to content from diverse formats delivered across numerous platforms.

File filtering plays an increasingly crucial role in markets as varied as eDiscovery, legal, security, and content management. It can be vital to any information-intensive organizations where data is key. In a litigious environment, for example, it's critical to be able to access all data, both structured and unstructured, for legal compliance. If the organization is unable to effectively filter through not only emails, but also the varied attachments in whatever formats they appear, it will be impossible to thoroughly monitor electronic communication for potential security breaches or data privacy loss.

A Comprehensive Offering

KeyView, an industry leading file filtering and file transformation solution, is a standalone solution, designed and built to work well with other components of Micro Focus IDOL. This includes our comprehensive set of PII, PHI, and PCI entity extraction grammars which reduce compliance and governance risk.

Quick View

- File detection, filtering and transformation with support for over 1,500 file formats
- No need for developers to stay current with native applications' releases
- Fast detection of file types without relying on filename extensions
- Full capture of text, metadata, and other content
- High-fidelity viewing, printing, and highlighting capabilities without the need to install original applications
- Simultaneous conversion of multiple documents with both in-process and out-of-process capabilities
- Filtering support for all major client-side and server-side platforms, including Windows, Linux (Intel, ARM), and macOS (Intel, Apple silicon)
- Fully maintained technology for greater reliability and support

Going beyond popular word processing, spreadsheet, and presentation formats, KeyView supports files embedded within files, such as email attachments, OLE objects in Office documents, and portfolio PDFs ensuring that all the information within a document set is extracted.

KeyView delivers powerful data extraction capabilities for extracting text, metadata, and other content while providing high-fidelity viewing and printing; it offers conversion to web-ready HTML or valid XML in order to maximize the data available for search and other critical business operations.

The high performance of KeyView, together with its low resource overhead requirement, means you can do more with less under many scenarios to increase the return on your data.

Key Benefits

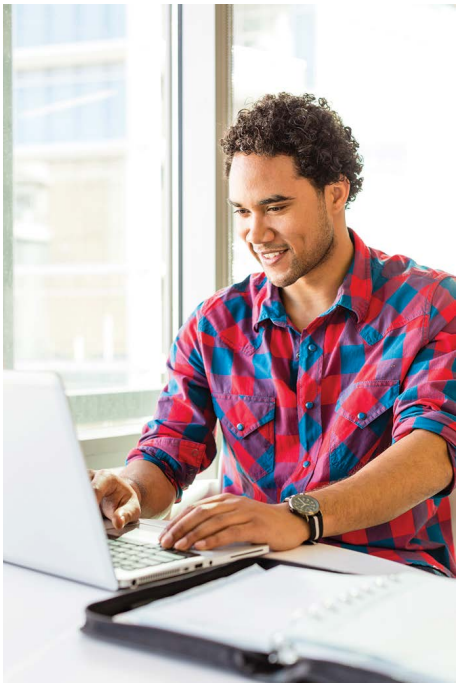
- **Reduce time to market**, engineering risk, internal development costs, and ongoing maintenance by leveraging Micro Focus's unparalleled history in file filtering and transformation.
- **Reduce information risk** of misprocessing crucial information or wasting valuable CPU time on irrelevant files. Instead of relying on falsifiable file name extensions or short magic numbers, KeyView forensically examines each file, focusing on the most differentiating characteristics first and going as deep as is needed to resolve ambiguity, resulting in faster answers and a lower error rate. KeyView goes beyond MIME type, clearly identifying files with non-existent or ambiguous MIME types (e.g., application/octet-stream), adding detail such as character set identification, encryption status, format classification and format version, giving you precision on which to base your downstream routing and processing decisions.
- **Don't miss a thing**, with the richest text extraction technology available. Get deep visibility into a wide variety of document formats, extracting content including body text, visible components (such as headers, footers, and section names), invisible text (such as notes, tracked changes, accessibility layers, and placeholder text), embedded, cached, and encapsulated objects, as well as file metadata (such as format specific fields, XMP, XrML, and classification).
- **Extend visibility** into scanned document sources using Optical Character Recognition (OCR) with a single configuration change for ultimate ease-of-deployment, adding visibility into digitally archived content with no code changes.
- **Achieve lower latency** by pipelining within documents, meaning that time-critical downstream processing can start earlier. This gives you results sooner, enabling you to meet tight SLAs.
- **Improve usability** by providing high-fidelity viewing (HTML5) capabilities for numerous document types including word processing, spreadsheet, presentation graphics, multimedia, and archive formats, all without the need to install native applications for each format.
- **Increase stability** by making use of KeyView's out-of-process capabilities for a session to continue in the event that a process fails.
- **Support your customers' journey** into emerging secure data storage trends by inspecting the content of files that have been protected by Microsoft's Azure Rights Management (RMS), part of AIP and MIP.
- **Maintain currency** with the ever-changing landscape of file formats as existing applications get updated and new ones are created. Your developers do not need to spend time keeping up-to-date with native application releases, past or present, and you don't need to rely on poorly maintained or unsupported software.

Key Features

Filter SDK

Integrate text extraction functionality into your own applications. The Filter SDK extracts text and metadata from a wide variety of file formats on numerous platforms and can automatically recognize over 1,500 document types. The Filter SDK supports both file-based and stream-based I/O operations, and provides in-process or out-of-process filtering.

- Thread-safe: filter multiple documents concurrently.
- File format detection without relying on file name extensions.
- Supported document classes: analytics, animation, CAD, database, desktop publishing, encapsulation, executable, font, GIS, library, movie, object module, outline, presentation graphics, raster image, schedule, scientific, sound, source code, spreadsheet, vector graphics and word processing.



- Deep filtering support: body text, endnotes, footnotes, and additional items such as document metadata are all included as part of the filtering process.
- Sample programs for faster implementation.
- Recursively extract files embedded within files, such as email attachments or embedded OLE objects.
- High resolution streaming for reduced latency.
- Partial filtering support, for reduced CPU usage.
- User configurable memory management.
- Redirected input and output streams, not restricted to file system access.
- Character encoding conversion.
- Source code identification.
- Optionally include OCR functionality.

HTML Export SDK

Convert virtually any document into high-fidelity HTML. Incorporating this technology into your web-based applications enables your end-users to access a document even if they do not have the appropriate plug-in or native application. With HTML Export, you control the content, structure, and format of the HTML output using either easily customized templates, or the flexible and robust APIs.

- Dynamically convert documents into web-ready compliant HTML.
- Create heading levels in the output file either by using the structure in the source document or by allowing Export to automatically generate a structure based on document properties, such as font or font attributes.
- Use call-backs to control aspects of the conversion process, such as file naming and the insertion of scripts.
- Insert predefined HTML markup at specific points in the output stream.
- Create navigable documents by automatically inserting links into target HTML.
- Break large documents into multiple linked pages for lower browser footprint.
- Apply Cascading Style Sheets (CSS) to improve output fidelity and align look and feel.
- Map paragraph and character styles in word processing documents to custom markup.
- Control the resolution of rasterized vector graphics to optimize storage requirements or image quality.
- Define the background, colors, and fonts used in the final HTML document, or maintain the source document's existing attributes.

PDF Export SDK

Create a PDF file out of virtually any document type. This SDK allows your end users to view files in PDF format, ensuring they can access the documents without having to download specific apps to display each file type.

XML Export SDK

Extract structure from virtually any document, creating well-formed, valid XML which is validated against a predefined Document Type Definition (DTD).

Structure identified in a source document includes metadata (title, author, subject…), document components (headers, footers, footnotes, endnotes, captions, bookmarks…), tagged text (chapters, sections, bulleted lists…), table components (sheet names, rows, columns, cell ranges…) and presentation components (notes, slide titles, slide descriptions…).

Additionally, Extensible Stylesheet Language (XSL) style sheets or Cascading Style Sheets (CSS) can be used to display the XML data in a human-friendly form.

Viewing SDK

Build high-fidelity document viewing capabilities into your native Windows applications. The Viewing SDK enables your users to open, view, and print virtually any document without having the native application or plug-in available. The SDK includes the following components:

- Viewing API (VAPI)—Windows messaging-based API
- Viewing .Net interface
- Sample programs

Contact us at:
www.microfocus.com

Like what you read? Share it.

Panopticon SDK

Identify and decrypt files that have been protected by Microsoft Azure Rights Management (RMS), part of Azure Information Protection (AIP), allowing your workflow to operate on the original, unencrypted file. You can use Panopticon with existing workflows to allow complete access to protected data for which the service has permission.

Learn more at
www.microfocus.com/keyview

Operating Environments

Supported programming languages

- C
- C++
- Java
- .NET

Supported platforms

- Windows/64
- Linux/x64
- Linux/AArch64
- macOS/x64
- macOS/M1
- Solaris/x64
- Solaris/SPARC
- AIX/Power

160-000384-002 | M | 05/21 | © 2021 Micro Focus or one of its affiliates. Micro Focus and the Micro Focus logo, among others, are trademarks or registered trademarks of Micro Focus or its subsidiaries or affiliated companies in the United Kingdom, United States and other countries. All other marks are the property of their respective owners.