DOCUMENT FILTERS | PRODUCT OVERVIEW

DOCUMENT FILTERS

DOCUMENT FILTERS IN ACTION

Identify all unstructured files and documents

Filter and extract every bit of content inside the files

Transform the extracted content into usable data

Output the data in the right format for your software

PDF · HTML · Image · High Definition Viewing · Multipage TIFF · Structured XML · Custom
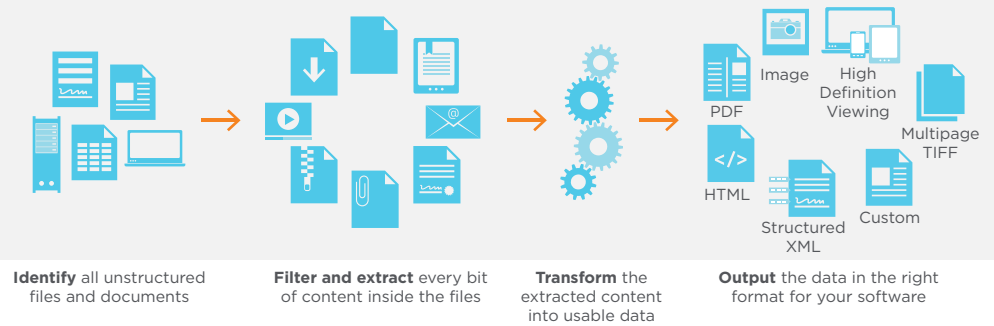
Hyland's Document Filters offering is a single SDK that gives software developers the power of deep inspection, format conversion, output manipulation and viewing for virtually any type of document. It also lets you work directly with extracted content in memory and choose the output modes that fit the diverse needs of your applications — TXT, HTML, structured XML, searchable PDF or image (JPG, PDF, PNG, BMP or TIFF).

Document Filters is the ideal OEM partner technology if your software must process unstructured content that it didn't create; enable near perfect viewing for different types of documents outside of native applications; perform a deep inspection of the text, metadata and hidden content of files; or convert, watermark and redact content into a unified format with pixel-level control.

For more than 25 years, Document Filters has powered some of the leading email archival, antivirus protection, content management, business intelligence, document imaging and intelligent capture software products on the market.

Document Filters also is the key catalyst driving content mining and intelligence gathering for applications such as compliance systems, eDiscovery, text analytics and Lucene deployments — offering a powerful and proven alternative to open-source options and other OEM solutions.

BENEFITS

Leading technology and innovation

- Transparent and proven performance
- More advanced and extensive roadmap
- Single platform and unified architecture

Partner-friendly licensing and contracts

- Unlimited distribution with no royalty fees
- Fixed licensing and support costs
- No restrictions for cloud deployment

Personalized service and support

- Accessible resources that respond quickly
- Simplified deployment with fewer issues
- Partners influence and advise direction

FEATURES

Deeply inspect content and data

- OCR document images and text, enabling full-text search
- Identify and extract every document, email, legacy, archive and container format you need — Word, Excel, PowerPoint, PDF, AutoCAD, ZIP, MSG, Visio and hundreds more
- Analyze all text and metadata in a file with deep-inspection capability that even uncovers previously hidden information, such as tracked changes, comments, notes, annotations and embedded web links
- Determine the true nature of content, ensuring that source information is accurately identified for filtering without relying on file-name extensions

View and annotate content

- Seamlessly manipulate and view content in high definition (HD) without the need for additional components such as ActiveX
- Eliminate the need for a third-party image manipulation package, applying precise redaction marks, annotations, Bates stamps and watermarks to content during output

Hyland®

Transform and output content

- Easily export content for further usage elsewhere by converting files into text, HTML, structured XML, paginated HTML, multipage TIFFs, images (JPG, BMP, PNG), searchable PDFs and your own proprietary formats
- Replicate original files through the Layout Engine that maps out exact, pixel-by-pixel coordinates of text, images and objects (instead of relying on simple character positioning)
- Render files at the page level and control the size of output and other variables, making it easy to create thumbnails or convert files with or without headers and footers
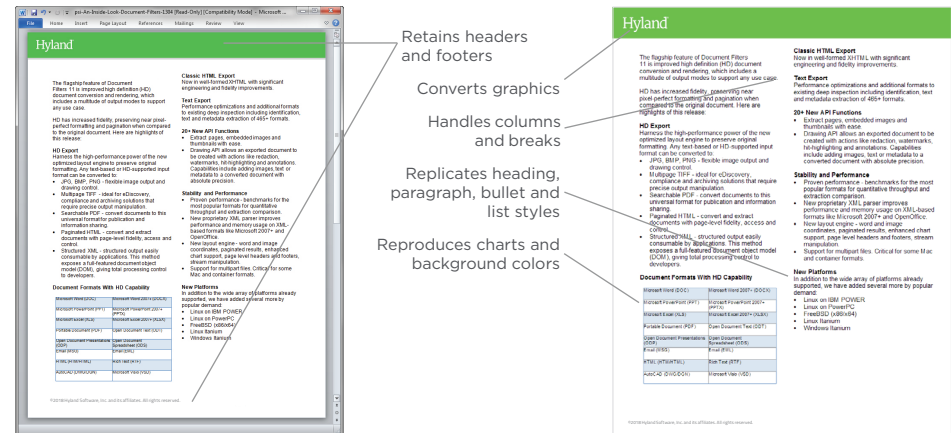
Leverage the most advanced search platform

- Deploy across 27 platforms including Windows, Mac OSX, Linux, Solaris, FreeBSD, iOS, Android, HP-UX and AIX — plus full support of character sets and encodings, such as Unicode
- Implement industry-leading extraction and throughput speeds, processing content faster with greater stability
- Embed Document Filters quickly and cost-effectively into your product with our flexible APIs for C, C++, COM, .NET, Python and Java
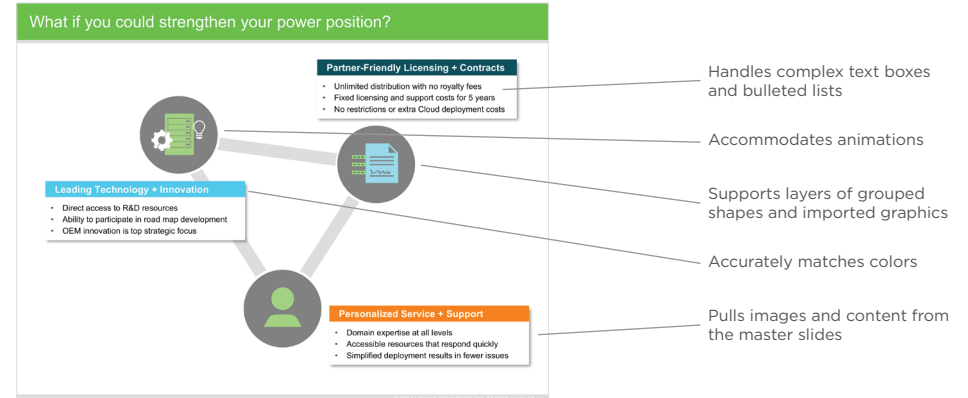
HD EXPORT EXAMPLES

The following images show a file in its native application and the high-definition image that Document Filters generates as a result of its file conversion and rendering process.

Microsoft Word

Retains headers and footers

Converts graphics

Handles columns and breaks

Replicates heading, paragraph, bullet and list styles

Reproduces charts and background colors

PowerPoint

Handles complex text boxes and bulleted lists

Accommodates animations

Supports layers of grouped shapes and imported graphics

Accurately matches colors

Pulls images and content from the master slides

Learn more at **Hyland.com/Document-Filters**

Hyland®