



SUPPORTED FORMATS – CV23

Table of Contents

- Archives 4
- CAD 5
- E-mails..... 6
- Multimedia 7
- Other Formats..... 8
- Presentations 9
- Rasters.....10
- Spreadsheets11
- Text & Markup.....12
- Vector Images.....13
- Word Processing14
- Supported MIME message formats16

Exalead CloudView converts file formats using a native conversion process. This document lists Exalead CloudView's ability to perform text and meta extraction, HTML preview, and thumbnail display for all Exalead CloudView-supported file formats.

This document is used by NETVIBES R&D to validate the file formats supported by Exalead CloudView. This validation is based on a set of files for each format, but given the wide range of possible variants and specificities, we cannot give any guarantee that a specific file format or a specific file will be handled properly. Document encoding is automatically detected by Exalead CloudView but better results are achieved with bigger files. Technical support on file formats is given on a best effort basis.

Archives

Text extraction for archives includes the archive listing but not its content.

Ext.	Description	Detected type
AR	Unix Static Library	application/x-ar
CPIO	Unix CPIO Archive	application/x-cpio
CPIO.GZ	Unix CPIO Archive	application/x-cpgz
CPIO.BZ2	Unix CPIO Archive	application/x-cpbz
OST	Outlook Offline Folder File	application/x-ms-outlook-folder
PST	Outlook Personal Information Store File	application/x-ms-outlook-folder
RAR	WinRAR Compressed Archive	application/x-rar
RPM	Red Hat Package Manager File	application/x-rpm
TAR	UNIX Tar	application/x-tar
TAR.BZ2	BZip2	application/x-tbz
TAR.GZ	GZip	application/x-tgz
ZIP	Zipped File	application/zip

Supported codecs are:

- BZIP2
- BZIP
- ZLIB
- LZX
- LZH

CAD

CAUTION: 3D CAD formats can be used with a OnePart product license only. These formats are not available with a standard Exalead CloudView license.

Ext.	Description	Detected type	Text & meta extraction	HTML preview	Thumbnail
DGN	Microstation v7 files (drawing files)	image/vnd.dgn	X	X	X
DGN	Microstation CAD V8	image/vnd.dgn	X	X	X
DWG	AutoCAD Drawing Database File	application/x-autocad	X	X	X
DXF	Drawing Exchange Format File	application/x-autocad	X	X	X
SLDASM	SolidWorks Assembly File	application/x-solidworks	X	X	X
SLDDRW	SolidWorks Drawing File	application/x-solidworks	X	X	X
SLDPRT	SolidWorks Part File	application/x-solidworks	X	X	X

E-mails

For more details on support for MIME-encoded formats, see “Supported MIME message formats”.

Ext.	Description	Detected type	Text & meta extraction	HTML preview	Thumbnail
DXL	Domino XML Language File	application/xml	X	X	X
EML	Microsoft Outlook Express	message/rfc822	X	X	X
MBOX	E-mail Mailbox File				
MSG	Microsoft Outlook Message	application/vnd.ms-outlook	X		
OST	See Archives				
PST	See Archives				

Multimedia

* Multimedia extraction only includes metadata (such as Author), File ID, and format information such as length, and bitrate.

Ext.	Description	Detected type	Text & meta extraction*	HTML preview	Thumbnail
ASF	ASF Video File	video/x-ms-asf	X		
AVI	AVI Multimedia file	video/msvideo	X	X	X
MOV	Quick Time Video File	video/quicktime	X		
MPEG	MPEG video formats	video/mpeg		X	X
MP3	Mp3 Audio File	audio/mp3	X		
MP4	MPEG video formats			X	X
OGG	OGG Audio File	audio/ogg	X		
QT	Quick Time Video File	video/quicktime	X		
SWF	Flash	application/x-shockwave-flash	X	X	X
TORRENT	Torrent File	application/x-torrent	X		
WAV	Windows Audio File	audio/x-wav	X		
WMA	Windows Media Audio File	video/x-ms-asf	X		
WMV	Windows Media Video File	video/x-ms-asf	X		

Other Formats

Ext.	Description	Detected type	Text & meta extraction	HTML preview	Thumbnail
CLASS	JAVA Class	application/java-vm	X		
DLL	Microsoft Windows DLL	application/octet-stream	X		
ELF	Linux binary File	application/x-linux-executable	X		
EXE	Microsoft Windows Executable	application/octet-stream	X		
JAR	JAVA Archive	application/java-archive	X		
MPP	Microsoft Project 2000/2003/2007	application/msproj	X	X	X
MPX	Microsoft Project 8	text/plain	text only	X	X
VCF	vCard File	text/plain	text only	X	X
VCS	vCalendar File	text/plain	text only	X	X

Presentations

Ext.	Description	Detected type	Text & meta extraction	HTML preview	Thumb nail
ODP	Open Office Presentation	application/vnd.oasis.opendocument.presentation	X	X	X
PPS	PowerPoint Slide Show	application/vnd.ms-powerpoint	X	X	X
PPT	Microsoft PowerPoint for Windows	application/vnd.ms-powerpoint	X	X	X
PPT	Microsoft PowerPoint for Windows 97	application/vnd.ms-powerpoint	X	X	X
PPTX	Microsoft PowerPoint 2007	application/vnd.openxmlformats-officedocument.presentationml.presentation	X	X	X
PPTX	Microsoft PowerPoint 2010	application/vnd.openxmlformats-officedocument.presentationml.presentation	X	X	X
PPTX	Microsoft PowerPoint 2013	application/vnd.openxmlformats-officedocument.presentationml.presentation	X	X	X
PPTX	Microsoft PowerPoint 2016	application/vnd.openxmlformats-officedocument.presentationml.presentation	X	X	X
SXI	StarOffice Impress Presentation	application/vnd.sun.xml.impress	X	X	X
WPD	WordPerfect X5 Presentation	application/wordperfect	X	X	X

Rasters

* Raster extraction only includes metadata (such as Author) and other format information such as image size, resolution, bit depth.

Ext.	Description	Detected type	Text & meta extraction	HTML preview	Thumbnail
BMP	Microsoft Windows Bitmap	image/bmp	X	X	X
BMP	OS/2 bitmap	application/octet-stream	X		
CGM	Computer Graphics Metafile	text/plain	X	X	X
GIF	Graphic Interchange Format	image/gif	X	X	X
JFIF	JPEG not in TIFF format	image/jpeg	X	X	X
JPEG	Progressive JPEG	image/jpeg	X	X	X
JPG	JPEG	image/jpeg	X	X	X
PNG	Portable Network Graphics	image/png	X	X	X
TIF	Image TIF	image/tiff	X	X	X
TIFF	Tagged Image File Format	image/tiff	X	X	X
XBM	X11 Bitmap Graphic	text/plain	X	X	X
XPM	X11 Pixmap Graphic	text/plain	X	X	X

Spreadsheets

Ext.	Description	Detected type	Text & meta extraction	HTML preview	Thumbnail
FCS	First Choice Spreadsheet	text/plain	text only	X	X
ODS	OpenOffice Spreadsheet	application/vnd.oasis.opendocument.spreadsheet	X	text only	
SXC	OpenOffice Calc	application/vnd.sun.xml.calc	X	X	X
SYLK	Microsoft Symbolic Link File	text/plain	X	text only	X
XLR	Works Spreadsheet	application/x-msoffice	text only	X	X
XLS	Microsoft Excel 95	application/vnd.ms-excel	X	X	X
XLS	Microsoft Excel 97-2003	application/vnd.ms-excel	X	X	X
XLS	Microsoft Excel XP	application/vnd.ms-excel	X	text only	
XLSX	Microsoft Excel 2007	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet	X	X	X
XLSX	Microsoft Excel 2010	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet	X	X	X
XLSX	Microsoft Excel 2013	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet	X	X	X
XLSX	Microsoft Excel 2016	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet	X	X	X
XLSM	Microsoft Excel 2007 and above	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet	X	X	X
WK3	Lotus 3 Worksheet	application/wordperfect	X	X	X
WK4	Lotus 4 Worksheet	application/wordperfect	X	X	X
WKS	Lotus 1-2-3 Spreadsheet	application/wordperfect	X	X	X
WPW	WP Works Word Processor File	text/plain	text only	X	X

Text & Markup

Ext.	Description	Detected type	Text & meta extraction	HTML preview	Thumbnail
HTM	Hypertext Markup Language Document	text/html	X	X	X
HTML	Hypertext Markup Language Document	text/html	text only	X	X
RTF	Rich Text Format	text/rtf	X	X	X
TXT	ANSI Text	text/plain	text only	X	X
TXT	ASCII Text	text/plain	text only	X	X
TXT	DOS Character Set	text/plain	text only	X	X
TXT	Unicode Text	text/plain	text only	X	X
TXT	UTF-8Text	text/plain	text only	X	X
WML	Wireless Markup Language File	text/html	text only	X	X
XHTML	XHTML Files	text/html	text only	X	X
XML	XML Files	application/xml	X	X	X

Vector Images

Ext.	Description	Detected type	Text & meta extraction	HTML preview	Thumbnail
AI	Adobe Illustrator	application/pdf	X	X	X
CDR	CorelDRAW	application/zip	X	X	
DGN	Microstation v7 files(drawing files)	image/vnd.dgn	X	X	X
EPS*	Encapsulated PostScript File	application/postscript	X	X	X
ODG	OpenOffice Draw	application/vnd.oasis.opendocument.graphics	X	X	X
PDF	Adobe PDF	application/pdf	X	X	X
PS*	PostScript File	application/postscript	X	X	X
PSD	Adobe Photoshop	image/x-photoshop	X	X	X
SXD	OpenOffice Draw	application/vnd.sun.xml.draw	X		
VSD	Microsoft Visio	application/vsd		X	X
VSDX	Microsoft Visio	application/vsd	text only	X	X

**ps2pdf must be installed*

Word Processing

Ext.	Description	Detected type	Text & meta extraction	HTML preview	Thumbnail
DOC	Microsoft Word for Windows 2000	application/msword	X	X	X
DOC	Microsoft Word for Windows 95	application/msword	X	X	
DOC	Microsoft Word for Windows 97	application/msword	X	X	X
DOC	Microsoft Word for Windows XP	application/msword	X		
DOCX	Microsoft Word for Office 2007	application/vnd.openxmlformats-officedocument.wordprocessingml.document	X	X	X
DOCX	Microsoft Word for Office 2010	application/vnd.openxmlformats-officedocument.wordprocessingml.document	X	X	X
DOCX	Microsoft Word for Office 2013	application/vnd.openxmlformats-officedocument.wordprocessingml.document	X	X	X
DOCX	Microsoft Word for Office 2016	application/vnd.openxmlformats-officedocument.wordprocessingml.document	X	X	X
DOCM	Microsoft Word for Office 2007 and above	application/vnd.openxmlformats-officedocument.wordprocessingml.document	X	X	X
MIF	Adobe FrameMaker	text/plain	text only	X	X
ODT	OpenOffice Open Document text	application/vnd.oasis.opendocument.text	X	X	X
SAM	Ami Pro Document	text/plain	text only	X	X
SXW	OpenOffice Writer Document	application/vnd.sun.xml.writer	X	X	X
WP	WordPerfect Document	application/wordperfect	X	X	X

	5.1, 5.2				
WRI	Windows Write Document	text/rtf	X	X	X
XY	XYWrite Document 3.5, 4.0	text/plain	text only	X	X

Supported MIME message formats

- MIME formats, including:
- EML
- MHTML (Messaging HTML)
- MHT (Web Archive)
- Simple Text Mail (defined in RFC 2822)
- TNEF Format
- MIME encodings, including:
- base64 (defined in RFC 1521)
- quoted-printable (defined in RFC 1521)
- utf-7 (defined in RFC 2152)
- binary (defined in RFC 1521)

Additionally, the body of a message can be encoded several ways. We support the following encodings:

- Text
- HTML
- RTF
- TNEF
- Text/rich text (defined in RFC1341)
- Embedded mail message (defined in RFC 822). This is handled as a link to a new message.

MIME message attachments can be stored in many formats. We process all attachments that are in Exalead CloudView-supported formats.