

KeyView

Software Version 12.3

HTML Export SDK Java Programming Guide



Document Release Date: June 2019
Software Release Date: June 2019

Legal notices

Copyright notice

© Copyright 2006-2019 Micro Focus or one of its affiliates.

The only warranties for products and services of Micro Focus and its affiliates and licensors ("Micro Focus") are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Micro Focus shall not be liable for technical or editorial errors or omissions contained herein. The information contained herein is subject to change without notice.

Documentation updates

The title page of this document contains the following identifying information:

- Software Version number, which indicates the software version.
- Document Release Date, which changes each time the document is updated.
- Software Release Date, which indicates the release date of this version of the software.

You can check for more recent versions of a document through the [MySupport portal](#). Many areas of the portal, including the one for documentation, require you to sign in with a Software Passport. If you need a Passport, you can create one when prompted to sign in.

Additionally, if you subscribe to the appropriate product support service, you will receive new or updated editions of documentation. Contact your Micro Focus sales representative for details.

Support

Visit the [MySupport portal](#) to access contact information and details about the products, services, and support that Micro Focus offers.

This portal also provides customer self-solve capabilities. It gives you a fast and efficient way to access interactive technical support tools needed to manage your business. As a valued support customer, you can benefit by using the MySupport portal to:

- Search for knowledge documents of interest
- Access product documentation
- View software vulnerability alerts
- Enter into discussions with other software customers
- Download software patches
- Manage software licenses, downloads, and support contracts
- Submit and track service requests
- Contact customer support
- View information about all services that Support offers

Many areas of the portal require you to sign in with a Software Passport. If you need a Passport, you can create one when prompted to sign in. To learn about the different access levels the portal uses, see the [Access Levels descriptions](#).

Contents

| | |
|--|----|
| Part I: Overview of HTML Export | 9 |
| Chapter 1: Introducing HTML Export | 11 |
| Overview | 11 |
| Features | 11 |
| Platforms, Compilers, and Dependencies | 12 |
| Supported Platforms | 12 |
| Supported Compilers | 13 |
| C++ Filter SDK | 14 |
| Software Dependencies | 14 |
| Windows Installation | 14 |
| UNIX Installation | 15 |
| Package Contents | 16 |
| License Information | 17 |
| Enable Advanced Document Readers | 17 |
| Update License Information | 18 |
| Directory Structure | 18 |
| Definition of Terms | 21 |
| Chapter 2: Getting Started | 22 |
| Architectural Overview | 22 |
| Enhance Performance | 24 |
| File Caching | 24 |
| Convert Files Out of Process | 24 |
| Configure Out-of-Process Conversions | 25 |
| Run Export Out of Process—Overview | 27 |
| Recommendations | 28 |
| Run Export Out of Process | 28 |
| Convert Files | 29 |
| Subfile Extraction | 30 |
| Convert Outlook Email without Using the Extraction API | 31 |
| Set Conversion Options | 31 |
| Set Conversion Options by Using the API | 31 |
| Set Conversion Options by Using the Template Files | 32 |
| Templates | 32 |
| Use the Export Demo Program | 33 |
| Change Input/Output Directories | 34 |
| Set Configuration Options | 35 |
| Convert PDF Files | 35 |
| Convert Rotated Text | 36 |
| Convert Files | 36 |
| Use the HTML Export API | 37 |
| Input/Output Operations | 38 |

| | |
|--|--------|
| Convert Files | 38 |
| Multithreaded Conversions | 41 |
| Use Callbacks | 42 |
| Example | 43 |
| Before Running Your Application | 43 |
| Export .NET Support | 43 |
| Third-Party Software and Redistributable Files | 43 |
| Sample Programs | 44 |
| Part II: Use the Export API | 45 |
| Chapter 3: Use the File Extraction API | 47 |
| Introduction | 47 |
| Extract Subfiles | 48 |
| Sanitize Absolute Paths | 49 |
| Extract Images | 50 |
| Recreate a File Hierarchy | 50 |
| Create a Root Node | 50 |
| Example | 51 |
| Extract Mail Metadata | 52 |
| Default Metadata Set | 52 |
| Extract the Default Metadata Set | 52 |
| Microsoft Outlook (MSG) Metadata | 53 |
| Extract MSG-Specific Metadata | 54 |
| Microsoft Outlook Express (EML) and Mailbox (MBX) Metadata | 54 |
| Extract EML- or MBX-Specific Metadata | 54 |
| Lotus Notes Database (NSF) Metadata | 55 |
| Extract NSF-Specific Metadata | 55 |
| Microsoft Personal Folders File (PST) Metadata | 55 |
| MAPI Properties | 55 |
| Extract PST-Specific Metadata | 56 |
| Exclude Metadata from the Extracted Text File | 57 |
| Extract Subfiles from Outlook Files | 57 |
| Extract Subfiles from Outlook Express Files | 57 |
| Extract Subfiles from Mailbox Files | 57 |
| Extract Subfiles from Outlook Personal Folders Files | 58 |
| Choose the Reader to use for PST Files | 58 |
| MAPI Attachment Methods | 60 |
| Open Secured PST Files | 60 |
| Detect PST Files While the Outlook Client is Running | 61 |
| Extract Subfiles from Lotus Domino XML Language Files | 61 |
| Extract .DXL Files to HTML | 62 |
| Extract Subfiles from Lotus Notes Database Files | 62 |
| System Requirements | 62 |
| Installation and Configuration | 63 |
| Windows | 63 |

| | |
|---|----|
| Solaris | 63 |
| AIX 5.x | 64 |
| Linux | 64 |
| Open Secured NSF Files | 65 |
| Format Note Subfiles | 65 |
| Extract Subfiles from PDF Files | 65 |
| Improve Performance for PDFs with Many Small Images | 65 |
| Extract Embedded OLE Objects | 65 |
| Extract Subfiles from ZIP Files | 66 |
| Default File Names for Extracted Subfiles | 66 |
| Default File Name for Mail Formats | 66 |
| Default File Name for Embedded OLE Objects | 67 |
| Exclude Japanese Guide Text | 68 |
| Chapter 4: Use the HTML Export API | 69 |
| Extract Metadata | 69 |
| Extract Metadata Using the API | 69 |
| Example | 70 |
| Extract Metadata Using a Template File | 72 |
| Examples | 72 |
| \$SUMMARYNN | 72 |
| \$SUMMARY | 73 |
| \$USERSUMMARY | 73 |
| Extract File Format Information | 74 |
| Example | 74 |
| Convert Character Sets | 75 |
| Determine the Character Set of the Output Text | 75 |
| Guidelines for Character Set Conversion | 76 |
| Examples of Character Set Conversion | 77 |
| Document Character Set Can be Determined | 77 |
| Document Character Set Cannot be Determined | 77 |
| Set the Character Set During Conversion | 78 |
| Set the Character Set During File Extraction from a Container | 78 |
| Map Styles | 79 |
| Use the Java API | 79 |
| Use a Template file | 80 |
| Use Style Sheets | 82 |
| Display Vector Graphics on UNIX and Linux | 83 |
| Search and Highlight Terms | 84 |
| Convert Revision Tracking Information | 84 |
| Extract Text from Text Boxes | 85 |
| Convert PDF Files | 85 |
| Use a Graphic-Based Reader | 86 |
| Use the kppdfdrdr Reader | 86 |
| Use the kppdf2rdr Reader | 86 |
| Specify the Graphic-based Reader | 87 |
| Convert PDF Files to Raster Images | 87 |

| | |
|--|-----|
| Convert PDF Files to a Logical Reading Order | 88 |
| Logical Reading Order and Paragraph Direction | 88 |
| Enable Logical Reading Order | 89 |
| Use the Java API | 89 |
| Use the formats_e.ini File | 90 |
| Generate a Table of Contents from PDF Bookmarks | 90 |
| Disable Bookmark Conversion | 91 |
| Convert Invisible Text | 91 |
| Toggle Invisible Text | 91 |
| Specify Opacity of Invisible Text | 91 |
| Convert Rotated Text | 92 |
| Control Hyphenation | 92 |
| Extract Custom Metadata from PDF Files | 93 |
| Convert Spreadsheet Files | 94 |
| Convert Hidden Text in Microsoft Excel Files | 94 |
| Convert Headers and Footers in Microsoft Excel 2003 Files | 94 |
| Specify Date and Time Format on UNIX Systems | 94 |
| Convert Very Large Numbers in Spreadsheet Cells to Precision Numbers | 95 |
| Extract Microsoft Excel Formulas | 95 |
| Convert XML Files | 97 |
| Configure Element Extraction for XML Documents | 97 |
| Modify Element Extraction Settings | 98 |
| Use the Java API | 98 |
| Use an Initialization File | 99 |
| Modify Element Extraction Settings in the kvxconfig.ini File | 99 |
| Specify an Element's Namespace and Attribute | 100 |
| Add Configuration Settings for Custom XML Document Types | 101 |
| Error Messages | 102 |
| Show Hidden Data | 105 |
| Hidden Data in Microsoft Documents | 105 |
| Toggle Word Comment Settings in the formats_e.ini File | 106 |
| Toggle PowerPoint Slide Note Settings in the formats_e.ini File | 106 |
| Exclude Japanese Guide Text | 107 |
| Source Code Identification | 107 |
| Chapter 5: Sample Programs | 109 |
| Introduction | 109 |
| ExtractExport | 109 |
| HtmlTest | 112 |
| HtmlConvFileToFile | 114 |
| Run HtmlConvFileToFile on Windows | 115 |
| Run HtmlConvFileToFile on UNIX | 116 |
| HtmlConvStreamToStream | 116 |
| Run HtmlConvStreamToStream on Windows | 117 |
| Run HtmlConvStreamToStream on UNIX | 117 |

| | |
|---|-----|
| Part III: Appendixes | 119 |
| Appendix A: Supported Formats | 121 |
| Supported Formats | 121 |
| Archive Formats | 122 |
| Binary Format | 125 |
| Computer-Aided Design Formats | 126 |
| Database Formats | 127 |
| Desktop Publishing | 128 |
| Display Formats | 128 |
| Graphic Formats | 129 |
| Mail Formats | 133 |
| Multimedia Formats | 136 |
| Presentation Formats | 139 |
| Spreadsheet Formats | 142 |
| Text and Markup Formats | 144 |
| Word Processing Formats | 145 |
| Appendix B: Detected Formats | 151 |
| Key to Detected Formats Table | 151 |
| Detected Formats | 153 |
| Appendix C: Character Sets | 202 |
| Multibyte and Bidirectional Support | 202 |
| Coded Character Sets | 210 |
| Appendix D: Extract and Format Lotus Notes Subfiles | 216 |
| Overview | 216 |
| Customize XML Templates | 216 |
| Use Demo Templates | 217 |
| Use Old Templates | 217 |
| Disable XML Templates | 217 |
| Template Elements and Attributes | 218 |
| Conditional Elements | 218 |
| Control Elements | 219 |
| Data Elements | 220 |
| Date and Time Formats | 223 |
| Lotus Notes Date and Time Formats | 223 |
| KeyView Date and Time Formats | 224 |
| Appendix E: Export Tokens | 229 |
| Appendix F: File Format Detection | 232 |
| Introduction | 232 |
| Extract Format Information | 232 |
| Determine Format Support | 232 |
| Refine Detection of Text Files | 233 |
| Change the Amount of File Data to Read | 233 |
| Change the Percentage of Allowed Non-ASCII Characters | 234 |
| Use the File Extension for Detection | 234 |

| | |
|---|---------|
| Allow Consecutive NULL Bytes in a Text File | 234 |
| Translate Format Information | 234 |
| Distinguish Between Formats | 235 |
| Determine a Document Reader | 236 |
| Category Values in formats_e.ini | 236 |
| Appendix G: Files Required for Redistribution | 240 |
| Core Files | 240 |
| Support Files | 241 |
| Document Readers and Writers | 243 |
| Appendix H: Password Protected Files | 250 |
| Supported Password Protected File Types | 250 |
| Export Password Protected Files | 251 |
| Open Password Protected Container Files | 251 |
| Send documentation feedback | 253 |

Part I: Overview of HTML Export

This section provides an overview of the Micro Focus IDOL KeyView Export SDK and describes how to use the Java implementation of the API.

- [Introducing HTML Export](#)
- [Getting Started](#)

Chapter 1: Introducing HTML Export

This guide is for developers who incorporate the Micro Focus KeyView HTML conversion technology into their custom web applications using a Java development environment. It is intended for readers who are familiar with HTML and Java.

This section describes the KeyView Export SDK package.

| | |
|--|----|
| • Overview | 11 |
| • Features | 11 |
| • Platforms, Compilers, and Dependencies | 12 |
| • Windows Installation | 14 |
| • UNIX Installation | 15 |
| • Package Contents | 16 |
| • License Information | 17 |
| • Directory Structure | 18 |
| • Definition of Terms | 21 |

Overview

HTML Export is part of the KeyView Export SDK. It enables you to convert virtually any document, spreadsheet, presentation, or graphic 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.

Export SDK supports a number of programming environments, such as Visual Basic, Java, .NET, and Delphi and runs on all popular operating system platforms including Windows, Solaris, HP-UX, IBM AIX, and Linux.

Export SDK is part of the KeyView suite of products. KeyView provides high-speed text extraction, conversion to web-ready HTML and well-formed XML, and high-fidelity document viewing.

Features

- Dynamically convert word processing, spreadsheet, presentation, and graphics files into web-ready, 4.0-compliant HTML.
- Export supports over 300 formats in 70 languages.
- Convert files either in-process or out of process. Out-of-process conversion ensures the stability and robustness of the calling application if a corrupt document causes an exception or causes the conversion process to fail.
- You can extract files embedded within files by using the File Extraction API, and then convert them

by using the Export API.

- Use redirected input/output. You can provide an input stream that is not restricted to file system access.
- Export automatically recognizes the file format being converted and uses the appropriate reader. Your application does not need to rely on file name extensions to determine the file format.
- 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 callbacks to control aspects of the conversion process, such as file naming and the insertion of scripts.
- Manage memory allocation to optimize speed and performance of application.
- Insert predefined HTML markup at specific points in the output stream.
- Create navigable documents by automatically inserting links into target HTML. You can also break large documents into multiple linked web pages.
- Apply Cascading Style Sheets (CSS) to improve the fidelity of the output.
- Map paragraph and character styles in word processing documents to any markup that you specify in the output.
- Control the resolution of rasterized vector graphics to optimize storage requirements or image quality.
- Select the target format for converted graphics, including GIF, JPEG, CGM, PNG, WMF, and Java on Windows, and Java and JPEG on Unix and Linux.
- Define the background, colors, and fonts used in the final HTML document, or maintain the source document's existing attributes.

Platforms, Compilers, and Dependencies

This section lists the supported platforms, supported compilers, and software dependencies for the KeyView software.

Supported Platforms

- CentOS 7
- FreeBSD 8.1 x86
- IBM AIX L6.1 PowerPC 32-bit and 64-bit
- IBM AIX L7.1 PowerPC 32-bit and 64-bit
- Mac OS X Mountain Lion 10.8 or higher on 32- and 64-bit Apple-Intel architecture
- Microsoft Windows Vista Business Edition x86 and x64. Other editions of Vista have not been tested, but are likely supported.

- Microsoft Windows 2008 Server Enterprise Edition x86 and x64
- Microsoft Windows 2008 Server R2
- Microsoft Windows 7 x86 and x64
- Microsoft Windows 8 x86 and x64
- Oracle Solaris 10 SPARC
- Oracle Solaris 10 x86 and x64
- Red Hat Enterprise Linux 5.0 x86 and x64
- Red Hat Enterprise Linux 6.0 x86 and x64
- SuSE Linux Enterprise Server 10, 10.1, 11, x86 and x64

Supported Compilers

| Platform | Architecture | Compiler Name | Compiler Version |
|-------------------|-------------------------------|---------------|---|
| Microsoft Windows | x86 | cl | Microsoft 32-bit C/C++ Optimizing Compiler Version 16.00.30319.01 for x86 |
| | x64 | cl | Microsoft C/C++ Optimizing Compiler Version 16.00.30319.01 for x64 |
| Sun Solaris | x86 64-bit | Sun Studio 12 | Sun C 5.9 SunOS_i386 Patch 124868-01 2007/07/12 |
| | SPARC 64-bit | Sun Studio 11 | Sun C 5.8 Patch 121015-06 2007/10/03 |
| Linux | x86 | gcc / g++ | 3.4.3 (Redhat 4), 4.1.0 (SuSE Linux 10) |
| | x64 | gcc / g++ | 4.1.0 (Redhat 4), 4.1.0 (SuSE Linux 10) |
| IBM AIX | Power | xLC_r / cc_r | IBM XL C/C++ Enterprise Edition V8.0 |
| Mac OSX | Apple-Intel 32-bit and 64-bit | LLVM | Apple LLVM 5.1 (clang-503.0.40) (based on LLVM 3.4svn) |
| FreeBSD | BSD x86 | gcc / g++ | 4.2.1 [FreeBSD] 20070719 |

Supported Compilers for Java Components

| Component | Compiler |
|-----------------|----------|
| Java components | Java 1.5 |

C++ Filter SDK

The C++ Filter SDK is supported on:

- Linux using GCC 5 or later
- Windows using Visual Studio 2015 or later

Software Dependencies

Some KeyView components require specific third-party software:

- Java Runtime Environment (JRE) or Java Software Developer Kit (JDK) version 1.5 is required for Java API and graphics conversion in Export SDK.
- Outlook 2002 or later is required to process Microsoft Outlook Personal Folders (PST) files using the MAPI-based reader (*pstsr*). The native PST readers (*pstxsr* and *pstnsr*) do not require Outlook.

NOTE:

You must install an edition of Microsoft Outlook (32-bit or 64-bit) that matches the KeyView software. For example, if you use 32-bit KeyView, install 32-bit Outlook. If you use 64-bit KeyView, install 64-bit Outlook.

If the editions do not match, KeyView returns `Error 32: KVErrror_PSTAccessFailed` and an error message from Microsoft Office Outlook is displayed: Either there is a no default mail client or the current mail client cannot fulfill the messaging request. Please run Microsoft Outlook and set it as the default mail client.

- Lotus Notes or Lotus Domino is required for Lotus Notes database (NSF) file processing. The minimum requirement is 6.5.1, but version 8.5 is recommended.
- The Microsoft .NET Framework is required if you are using the .NET implementation of the API.
- Microsoft Visual C++ 2013 and Microsoft Visual C++ 2010 Redistributables (Windows only).

Windows Installation

To install the SDK on Windows, use the following procedure.

To install the SDK

1. Run the installation program, *KeyViewProductNameSDK_VersionNumber_OS.exe*, where *ProductName* is the name of the product, *VersionNumber* is the product version number, and *OS* is the operating system.

For example:

`KeyViewExportSDK_12.3_Windows_X86_64.exe`


The installation wizard opens.

2. Read the instructions and click **Next**.

The License Agreement page opens.

3. Read the agreement. If you agree to the terms, click **I accept the agreement**, and then click **Next**.

The Installation Directory page opens.

4. Select the directory in which to install the SDK. To specify a directory other than the default, click , and then specify another directory. After choosing where to install the SDK, click **Next**.

The License Key page opens.

5. Type the company name and license key that were provided when you purchased KeyView, and then click **Next**.
 - The company name is case sensitive.
 - The license key is a string that contains 31 characters.

NOTE:

The installation program validates the company name and license key and generates the file `install\OS\bin\kv.lic` (where `install` is your chosen installation folder and `OS` is the name of the operating system platform). The license information is validated when the KeyView API is used. If you do not enter a license key at this step, or if you enter invalid information, the KeyView SDK is installed, but the API does not function. When you obtain a valid license key, you can either re-install the KeyView SDK, or manually update the license key file (`kv.lic`) with the new information. For more information, see [License Information, on page 17](#).

The Pre-Installation Summary dialog box opens.

6. Review the settings, and then click **Next**.

The SDK is installed.

7. Click **Finish**.

UNIX Installation

To install the SDK, use one of the following procedures.

To install the SDK from the graphical interface

- Run the installation program and follow the on-screen instructions.

To install the SDK from the console

1. Run the installation program from the console as follows:

```
./KeyViewExportSDK_VersionNumber_Platform.exe --mode text
```

where:

VersionNumber is the product version.

Platform is the name of the platform.

2. Read the welcome message and instructions and press `Enter`.

The first page of the license agreement is displayed.

3. Read the license information, pressing `Enter` to continue through the text. After you finish reading the text, and if you accept the agreement, type `y` and press `Enter`.

You are asked to choose an installation folder.

4. Type an absolute path or press `Enter` to accept the default location.

You are asked for license information.

5. At the **Company Name** prompt, type the company name that was provided when you purchased KeyView, and then press `Enter`. The company name is case sensitive.
6. At the **License Key** prompt, type the license key that was provided when you purchased KeyView, and then press `Enter`. The license key is a string that contains 31 characters.

NOTE:

The installation program generates the file `install\OS\bin\kv.lic` (where `install` is your chosen installation folder and `OS` is the name of the operating system platform). The license information is validated when the KeyView API is used. If you do not enter a license key at this step, or if you enter invalid information, the KeyView SDK is installed but the API does not function. When you obtain a valid license key, you can either re-install the KeyView SDK, or manually update the license key file (`kv.lic`) with the new information. For more information, see [License Information, on the next page](#).

The Pre-Installation summary is displayed.

7. If you are satisfied with the information displayed in the summary, press `Enter`.

The SDK is installed.

Package Contents

The Export installation contains:

- Libraries and executable files necessary for converting source documents into high-quality, web-ready HTML (see [Files Required for Redistribution, on page 240](#)).
- The include files that define the functions and structures used by the application to establish an interface with Export:

`adinfo.h`

`kverrorcodes.h`

`kvhtml.h`

`kvtypes.h`
`kvxtract.h`

- The Java API implemented in the `com.verity.api.export` package contained in the `KeyView.jar` file.
- Several sample programs that demonstrate Export's functionality.
- Sample images that can be used as navigation buttons and background textures in your output.
- Template files that enable you to set conversion options without modifying at the API level. They can be used to generate a wide range of output, from highly-stylized user-defined HTML to stripped-down, text-only output suitable for use with an indexing engine.
- Sample style sheet: `WordStyle.css` (for word processing documents).

License Information

During installation, the installation program validates the organization name and license key that you enter, and generates the `install/OS/bin/kv.lic` file, where `install` is the directory in which you installed KeyView, and `OS` is the operating system. This file is opened and validated when the KeyView API is used.

The `kv.lic` file contains the organization name and the 31-digit license key you specified during installation. The contents of a `kv.lic` file looks similar to the following:

```
Company Name  
XXXXXXXX-XXXXXXXX-XXXXXXXX-XXXXXXXX
```

The license key controls whether the following are enabled:

- the full version of the KeyView SDK
- the trial version of the KeyView SDK
- language detection and advanced document readers—The following components are considered advanced features, and are licensed separately:
 - Microsoft Outlook Personal Folders (PST) readers (`pstsr`, `pstnsr`, and `pstxsr`)
 - Lotus Notes database (NSF) reader (`nsfsr`)
 - Mailbox (MBX) reader (`mbxsr`)
 - Character set detection library (`kvlangdetect`)

If you change the license key at any time, you must update the licensing information in the `kv.lic` file. See [Update License Information](#).

Enable Advanced Document Readers

To enable advanced readers in one of the KeyView SDKs, you must obtain an appropriate license key from Micro Focus and update the installed license key with the new information as described in [Update License Information](#).

If you are enabling the MBX reader in an existing installation of Export, in addition to updating the license key, change the parameter `208=eml` to `208=mbx` in the `formats_e.ini` file.

Update License Information

If you currently have an evaluation version of KeyView and have purchased a full version of the SDK, or you are adding a document reader (for example, the PST reader), you must update the license information that was installed with the original version of the KeyView SDK.

If you installed a full version of KeyView, but did not enter licensing information at the time of installation, you must also update the license information.

To update the information, do one of the following:

- Manually update the license information that is stored in the text file named `kv.lic`.
- Re-install the product and enter the new license information when prompted.

To update the KeyView license information

1. Open the license key file, `kv.lic`, in a text editor. The file is in the `install\OS\bin` directory, where `install` is the directory in which you installed KeyView, and `OS` is the operating system. The file contains the following text:

```
COMPANY NAME  
XXXXXXX-XXXXXXX-XXXXXXX-XXXXXXX
```

2. Replace the text `COMPANY NAME` with the company name that appears at the top of the License Key Sheet provided by Micro Focus. Enter the text exactly as it appears in the document.
3. Replace the characters `XXXXXX-XXXXXXX-XXXXXXX-XXXXXXX` with the appropriate license key from the License Key Sheet provided by Micro Focus. The license key is listed in the **Key** column in the **Standalone Products** table. The key is a string that contains 31 characters, for example, `2TQD22D-2M6FV66-2KPF23S-2GEM5AB`. Enter the characters exactly as they appear in the document, including the dashes, but do not include a leading or trailing space.
4. The finished `kv.lic` file looks similar to the following:

```
Autonomy  
24QD22D-2M6FV66-2KPF23S-2G8M59B
```

5. Save the `kv.lic` file.

Directory Structure

The following table describes the directories created during the HTML Export installation. The variable `install` is the path name of the Export installation directory (for example, `/usr/autonomy/KeyviewExportSDK` on UNIX, or `C:\Program Files\Autonomy\KeyviewExportSDK` on Windows). On UNIX, the HTML Export directory is named `/htmlxpt`.

The variable `OS` is the operating system for which the SDK is installed. For example, the `bin` directory on a standard 32-bit Windows installation would be located at `C:\Program Files\Autonomy\KeyviewExportSDK\WINDOWS\bin`.

HTML Export installed directory structure

| Directory | Description |
|--|---|
| <i>install\OS\bin</i> | Contains the libraries, executables for sample programs, the Java program (<i>kvraster.class</i>), the Java applet (<i>kvvector.jar</i>), the format detection file, <i>formats_e.ini</i> , the license key file (<i>kv.lic</i>), and a number of other supporting files. |
| <i>install\OS\lib</i> | (Solaris installations only) Contains the redistributable <i>libstlport.so.1</i> library, which is required to run KeyView on Solaris platforms. |
| <i>install\dotnet\sample</i> | The <i>HtmlConvFileToFile.cs</i> C# sample program demonstrating the .NET interface. |
| <i>install\HTML Export\docs</i> | Contains the converted version of the sample word processing, spreadsheet, and presentations files. |
| <i>install\HTML Export\guide</i> | Contains the <i>HTML Export C and COM Programming Guide</i> and <i>HTML Export Java Programming Guide</i> in HTML and PDF format. |
| <i>install\HTML Export\include</i> | Contains the header files. |
| <i>install\HTML Export\programs\bin</i> | Contains the executable files for the Visual Basic sample program called Export Demo. |
| <i>install\HTML Export\programs\callback</i> | Contains the C source code and supporting files for a sample program that demonstrates how user callbacks can dynamically shape the HTML conversion. |
| <i>install\HTML Export\programs\cnv2html</i> | Contains the C source code for a sample program that creates a single HTML file. The executable for this sample program is in the <i>bin</i> directory. |
| <i>install\HTML Export\programs\cnv2htmlloop</i> | Contains the C source code for a sample program that creates a single HTML file out of process. |
| <i>install\HTML Export\programs\comsamp</i> | Contains the Visual Basic source code and supporting files for a sample program that demonstrates the use of the COM interface. |
| <i>install\HTML Export\programs\ExportDemo</i> | Contains the source code for a sample Visual Basic program. The executable for this program is in the <i>bin</i> directory. The Export Demo is available through the Start menu. |
| <i>install\HTML Export\programs\htmlini</i> | Contains the C source code and supporting files for a sample program that uses template files to set the conversion options. |
| <i>install\HTML Export\programs\htmlloop</i> | Contains a Visual C++ sample program that uses the Microsoft Foundation Classes (MFC) to provide out-of-process HTML |

HTML Export installed directory structure, continued

| Directory | Description |
|---|---|
| | conversion using the COM automation server. |
| <i>install</i> \HTML Export\programs\images | Contains the background graphics and navigation buttons used by the template files. |
| <i>install</i> \HTML Export\programs\index | Contains the C source code and supporting files for a sample program that produces text-only HTML. |
| <i>install</i> \HTML Export\programs\ini | Contains the template files used to set the conversion options in the C API. |
| <i>install</i> \HTML Export\programs\io_samp | Contains the C source code and supporting files for a sample program that demonstrates how to input a stream by providing a simple wrapper around the ANSI C interface <code>fread()</code> , <code>fopen()</code> , and so on. |
| <i>install</i> \HTML Export\programs\jstree | Contains the C source code and supporting files for a sample program that employs JavaScript to produce an expandable table of contents in a frame-based HTML output file. |
| <i>install</i> \HTML Export\programs\jvtree | Contains the C source code and supporting files for a sample program that uses the <code>JVTree</code> Java applet in creating an expandable table of contents in a frame-based HTML output file. |
| <i>install</i> \HTML Export\programs\jvtree_ demo | C code sample that creates a frame-based HTML stream using the <code>JVTree</code> Java applet to display the table of contents. |
| <i>install</i> \HTML Export\programs\pdfini | Contains the configuration file used to extract custom metadata from PDF documents. |
| <i>install</i> \HTML Export\programs\tempout | The default output directory for converted files. Contains a sample style sheet. |
| <i>install</i> \HTML Export\programs\ tstextract | Contains the C source code and supporting files for a sample program that demonstrates the File Extraction interface. |
| <i>install</i> \HTML Export\programs\wizard | Contains the source code and supporting files for the Visual Basic program HTML Export Template Wizard. |
| <i>install</i> \HTML Export\rel_ notes | Contains the <i>HTML Export Release Notes</i> in HTML and PDF format. |
| <i>install</i> \javaapi\ini | Contains the template files used with the Java API. |
| <i>install</i> \javaapi\javadoc | Contains the Javadoc for the Java API. |

HTML Export installed directory structure, continued

| Directory | Description |
|-------------------------------|--|
| <i>install\javaapi\sample</i> | Contains the source files and sample programs for the Java API. |
| <i>install\testdocs</i> | Contains sample word processing, spreadsheet, and presentation files that you can use to test HTML Export's options. You might also find this directory useful when testing your own applications. |

Definition of Terms

The following are specialized terms used throughout the guide.

| | |
|-------------------------------|--|
| anchor | HTML markup that defines both anchors and hyperlinks. An anchor is a named place in a document to which other documents can form a link. Anchors use the HTML anchor tags (<code><a></code> <code></code>) to facilitate navigation within a document. |
| block | All source document content (including subheadings) associated with Heading Level 1. Export identifies and/or generates blocks from the input stream for the implementation of the your HTML markup. |
| block chunk or chunk | All source document content associated with Heading Levels 2 through 6. Chunks are subdivisions of blocks. You can supply specific HTML markup for the different levels of block chunks. |
| callback | A function optionally supplied by your application and called from the Export API. For example, callbacks allow your application to monitor the progress of the conversion process dynamically. |
| stream | Transmission of a file's content between memory and disk in a continuous flow. |
| token | The vehicle for conveying specific types of information to and from the API during the conversion process. Tokens are placeholders for markup that appears in the output. See Export Tokens, on page 229 . |

Chapter 2: Getting Started

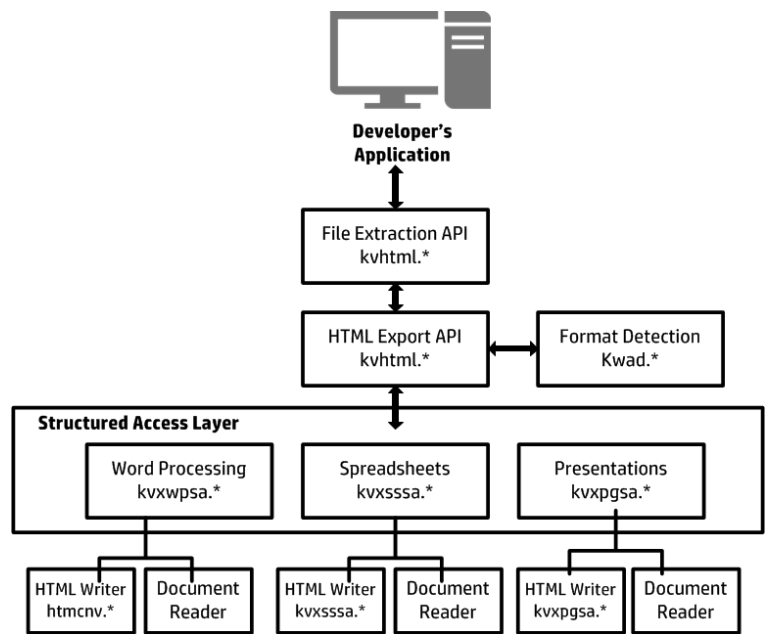
This section provides an overview of the HTML Export SDK and describes how to use the Java implementation of the API. It contains the following topics:

- [Architectural Overview](#) 22
- [Enhance Performance](#) 24
- [Convert Files Out of Process](#) 24
- [Convert Files](#) 29
- [Subfile Extraction](#) 30
- [Set Conversion Options](#) 31
- [Use the Export Demo Program](#) 33
- [Use the HTML Export API](#) 37
- [Export .NET Support](#) 43

Architectural Overview

The general architecture of the KeyView HTML conversion technology is the same across all supported platforms and is illustrated in the following diagram.

HTML Export Architecture



Each component is described in the following table.

Architectural Components

| Component | Description |
|-------------------------|--|
| Developer's Application | The developer's application interfaces directly with the HTML Export API through either a C, Java, or COM implementation. |
| File Extraction API | The File Extraction API opens a file and extracts the file's subfiles so that they are available for conversion. See Use the File Extraction API, on page 47 . |
| HTML Export API | The HTML Export API exposes the functionality of HTML Export and controls all other HTML Export modules during the conversion process. |
| Format Detection Module | The format detection module determines the file type of the source file, which enables the HTML Export interface to load the appropriate structured access layer module and document reader. See File Format Detection, on page 232 . |
| Structured Access Layer | <p>The structured access layer contains three modules: one for word processing, one for spreadsheets, and one for presentations and graphics. Information from the format detection module determines which access layer module operates at this stage of the conversion. The structured access layer performs the following:</p> <ol style="list-style-type: none">1. Loads the appropriate document reader.2. Processes the data stream from the document reader.3. Determines table of contents entries.4. Sends the stream to the appropriate HTML writer.5. Accepts the HTML stream from the HTML writer.6. Generates the HTML output file with a table of contents, metadata, and the document's contents, and sends it to the HTML Export interface. |
| Document Reader | Each document reader reads a specific file format and sends a text stream of the document to the structured access layer. Word processing readers return a <i>token stream</i> to the structured access layer. A token stream contains the document contents and messages (tokens) that precede the content and identify the type of information that follows them. Each reader is loaded as required by the structured access layer. See Document Readers and Writers, on page 243 for a complete list of document readers. |
| HTML Writers | Each HTML writer accepts a text stream or token stream from the structured access layer and generates an equivalent HTML stream that is sent back to the structured access layer. The structured access layer then generates the output file. See Document Readers and Writers, on page 243 for a list of format writers. |

Enhance Performance

KeyView is designed for optimal performance out of the box. However, there are some parameters that can be adjusted to improve performance specifically for your system.

File Caching

To reduce the frequency of I/O operations, and consequently improve performance, the KeyView readers load file data into memory. The readers then read the data from the cache rather than from the physical disk. You can configure the amount of memory used for file caching through the `formats_e.ini` file. Generally, when you increase the memory, performance improves.

By default, KeyView uses a maximum of 1 MB of memory for each thread. If the file data is larger than 1 MB, up to 1 MB of data is cached and the data beyond 1 MB is read from disk. The minimum amount of memory that can be used for file caching is 64 KB.

To determine a reasonable value, divide the maximum amount of memory that you want KeyView to use for file caching by the total number of threads. For example, if you want KeyView to use a maximum of 50 MB of memory and have 10 threads, set the value to 5 MB.

To modify the memory allocated for file caching, change the value for the following parameter in the `[DiskCache]` section of the `formats_e.ini` file:

```
DiskCacheSize=1024
```

The value is in kilobytes. If you do not set this parameter or if you set it to 0 (zero), the minimum value of 64 KB is used.

The `formats_e.ini` file is in the directory `install\OS\bin`, where `install` is the path name of the Export installation directory and `OS` is the name of the operating system.

Convert Files Out of Process

Export can run independently from the calling application. This is called *out of process*. Out-of-process conversions protect the stability of the calling application in the rare case when a malformed document causes Export to fail. You can also run Export in the same process as the calling application. This is called *in-process*. However, it is strongly recommended you convert documents out of process whenever possible.

The Export out-of-process framework uses a client-server architecture. The calling application sends an out-of-process conversion request to the Service Request Broker in the main Export process. The Broker then creates, monitors, and manages a Servant process for the request—each request is handled by one independent Servant process. Data is exchanged between the application thread and the Servant through TCP/IP sockets. The source data is sent to the Servant process as a data stream or file, converted in the Servant, and then returned to the application thread. At that point, the application can either terminate the Servant process or send more data for conversion.

Multiple conversion requests can be sent from multiple threads in the calling application simultaneously. All requests sent from one thread are processed by the Servant mapped to that thread; in other words, each thread can only have one Servant to process its conversion requests.

Any standard conversion errors generated by the Servant are sent to the application.

NOTE:

Currently, the main Export process and Servant processes must run on the same host.

The following are requirements for running Export out of process:

- Internet Protocol (TCP/IP) must be installed
- Multithreaded processing must be supported on the operating system platform
- The user application must be built with a multithreaded runtime library

The following methods run in-process or out of process:

- `convert`
- `convertTo`
- `getSummaryInfo`

NOTE: When converting out of process, these methods must be called after the call to start an out-of-process session and before the call to end an out-of-process session.

Other HTML Export methods and the File Extraction methods always run in-process.

Configure Out-of-Process Conversions

Although most components of the out-of-process conversion are transparent, the following parameters are configurable:

- File-size threshold/temporary file location
- Conversion time-out
- Listener port numbers and time-out
- Connection time-out and retry
- Servant process name

These parameters are defined internally, but you can override the default by defining the parameter in the `formats_e.ini` file. The `formats_e.ini` file is in the directory `install\OS\bin`, where `install` is the path name of the Export installation directory and `OS` is the name of the operating system.

To set the parameters, add the following section to the `formats_e.ini` file:

```
[KVExportOOPOptions]
TempFileSizeMark=
TempFilePath=
WaitForConvert=
WaitForConnectionTime=
ListenerPortList=

ConnectRetryInterval=
ConnectRetry=
ServantName=
```

Each parameter is described in the following table.

The default values for these parameters are set to ensure reasonable performance on most systems. If you are processing a large number of files, or running Export on a slow machine, you might need to increase some of the time-out and retry values.

Parameters for Out-of-Process Conversion

| Parameter | Description |
|--|--|
| TempFileSizeMark unit = megabytes default=10 | The <i>file-size threshold</i> . If the input file received by the Servant is larger than this value, temporary files are created to store the data. The directory in which the temporary files are stored is defined by the TempFilePath parameter. If the file received is smaller than this value, the data is stored in memory in the Servant. This only applies when the input is a stream. |
| TempFilePath type = file path default = current working directory | The directory in which temporary files are stored. Temporary files are created when the input file surpasses the file-size threshold (TempFileSizeMark). If the Servant cannot access the file path, an error is generated. This only applies when converting in stream mode. |
| WaitForConvert unit = seconds default = 1800 range = 30~3600 | The length of time to wait for a Servant to convert a file. If the conversion is not completed within the specified time, the error message "wait for child process failed" is generated. |
| WaitForConnectionTime unit = seconds default = 180 range = 15~600 | The length of time to wait for the Servant to connect to the application thread after the application has sent a conversion request to the Broker. If the Servant does not connect within the specified time, the error message "wait for child process failed" is generated. If there are many Servant processes running simultaneously, you might need to increase this value. |
| ListenerPortList type = integer default = 9985, 9986, 9987, 9988, 9989 | The TCP/IP port number(s) used for communication between the calling application and the Servant. You can specify a single port number or a series of numbers (enter the number separated by commas). |
| ListenerTimeout unit = seconds default = 10 range = 5~30 | The length of time to wait for the Servant listener thread to get a process ID from the Servant after the connection is established. If the ID is not obtained within the specified time, the error message "wait for child process failed" is generated. During this time, no other Servant can connect with the application. |
| ConnectRetryInterval unit = microseconds | The length of time to wait after a Servant has failed to connect to the application before it retries the connection. A Servant might be unable to |

Parameters for Out-of-Process Conversion, continued

| Parameter | Description |
|---|--|
| default = 0.1 range = 50000~500000 | connect because the application is waiting for another Servant to send a process ID. To calculate the <i>total retry interval</i> , the value set here is added to the platform-specific TCP retry value (on Windows, this is 1 second). |
| ConnectRetry type = integer default = 120 range = 30~600 | The number of attempts the Servant makes to connect to the calling application. This value and the total retry interval determine the total delay time. The total delay is calculated as follows: $\text{ConnectRetryInterval} + \text{platform-specific_TCP_retry_value} * \text{ConnectRetry}$ For example, if the <code>ConnectRetryInterval</code> is set to 2 seconds, and the Export process is running on Windows (the default TCP retry value on Windows is 1 second), the total delay would be: $2 + 1 * 120 = 360$ The Servant would attempt to connect to the application every 3 seconds for 120 attempts for a total of 360 seconds. |
| ServantName type = string default = servant | The name of the Servant process. To move the Servant to another location, enter a fully qualified path. |

Run Export Out of Process—Overview

To convert files out of process

1. If required, set parameters for the out-of-process conversion in the `formats_e.ini` file. See [Configure Out-of-Process Conversions, on page 25](#).
2. Instantiate an `HtmlExport` object.
3. Define the conversion options.
4. Initialize an out-of-process session.
5. Convert the input and/or call other methods that can run out of process.
6. Shutdown the out-of-process session.
7. Repeat [Step 3](#) to [Step 6](#) for additional files.
8. Terminate the out-of-process session and the Servant process.
9. Shutdown the Export session.

Recommendations

- To ensure multithreaded conversions are thread-safe, you must create a unique context pointer for every thread by instantiating an `HtmlExport` object. In addition, threads must not share context pointers, and the same context pointer must be used for all API calls in the same thread. Creating a context pointer for every thread does not affect performance because the context pointer uses minimal resources.
- All methods that can run out of process must be called within the out-of-process session, that is, after the call to initialize the out-of-process session and before the call to end the out-of-process session.
- When terminating an out-of-process session, persist the Servant process by setting the Boolean flag `bKeepServantAlive` in the `endOOPSession` method. If the Servant process remains active, subsequent conversion requests are processed more quickly because the Servant process is already prepared to receive data. Only terminate the Servant when there are no more out-of-process requests.
- To recover from a failure in the Servant process, start a new out-of-process session. This creates a new Servant process for the next conversion.

Run Export Out of Process

The `HtmlTest.java` sample program demonstrates how to run Export out of process.

To convert files out of process in the Java API

1. If required, set parameters for the out-of-process conversion in the `formats_e.ini` file. See [Configure Out-of-Process Conversions, on page 25](#).
2. Create an instance of the `HtmlExport` class.

```
HtmlExport objHtmlExport = new HtmlExport();
```
3. If you are using a template file to set the conversion options, call the `setIniFileName` method.

```
objHtmlExport.setIniFileName(iniFile);
```
4. If you are using the API to set the conversion options, create instances of the following classes:

```
HtmlOptionInfo  
HtmlTemplateInfo  
HtmlTOCOptionInfo  
StyleMapping  
HtmlHeadingInfo
```

Set the classes to the current `HtmlExport` object using the appropriate set methods. If you do not set the classes before calling the `startOOPSession` method, default values are used.
5. Set the location of the Export libraries by calling the `setExportDirectory` method. These are

normally in the directory `install\OS\bin`, where `install` is the path name of the Export installation directory and `OS` is the name of the operating system.

```
objHtmlExport.setExportDirectory(exportDir);
```

6. Optionally, set the input source as either a file or stream by calling the `setInputSource` method.

```
//set the input source as file
```

```
objHtmlExport.setInputSource(inFile);
```

```
//set the input source as stream
```

```
inf = new File(inFile);
```

```
fis = new FileInputStream(inf);
```

```
objHtmlExport.setInputSource(fis);
```

7. Set up an out-of-process session by calling the `startOOPSession` method. This initializes the out-of-process session, creates a Servant process, establishes a communication channel between the application thread and the Servant, and sends the data to the Servant.

```
objHtmlExport.startOOPSession();
```

8. Convert the input and generate the output files by calling the `convertTo` method. You cannot use the `convert` methods that set the input source because the input source must be set *before* the out-of-process session is initialized. The `convertTo` method can only be called once in a single out-of-process session.

```
objHtmlExport.convertTo(outFile);
```

9. Terminate the out-of-process session by calling the `endOOPSession` method. The Servant ends the current conversion session, and releases the source data and session resources.

```
objHtmlExport.endOOPSession(TRUE);
```

10. Repeat [Step 3](#) through [Step 9](#) for additional files.

11. After all files are converted, terminate the out-of-process session *and* the Servant process by calling `endOOPSession` and setting the Boolean to `FALSE`.

```
objHtmlExport.endOOPSession(FALSE);
```

12. Terminate the Export session and free allocated system resources by calling the `shutdownExport()` method.

```
objHtmlExport.shutdownExport();
```

Convert Files

KeyView Export SDK enables you to *convert* many different types of documents to HTML.

Converting is the process of extracting the text from a document without the application-specific markup, and applying HTML markup. The conversion process can also include the following:

- Extracting subfiles to expose all subfiles for conversion. See [Subfile Extraction, below](#).
- Setting conversion options to determine the content, structure, and appearance of the HTML output. See [Set Conversion Options, on the next page](#).
- Extracting the file's format to detect a file's format and report the information to the API, which in turn reports the information to the developer's application. See [Extract File Format Information, on page 74](#).
- Extracting selected metadata (document properties) from a file. See [Extract Metadata, on page 69](#).
- Converting character sets to control the character set of both the input and the output text. See [Convert Character Sets, on page 75](#).
- Implementing callbacks to control the conversion while it is in progress. See [Use Callbacks, on page 42](#).

You can use one of the following methods to convert documents:

- Use the Export Demo sample program. This Visual Basic program demonstrates most Export API functionality and is the easiest way to get started. See [Use the Export Demo Program, on page 33](#).
- Use the Java implementation of the API. See [Use the HTML Export API, on page 37](#) and the Javadoc in the directory `install\javaapi\javadoc`, where `install` is the path name of the Export installation directory.
- Use the Java sample programs. See [Sample Programs, on page 109](#).

NOTE:

Micro Focus strongly recommends that you convert documents *out of process*. During out-of-process conversion, Export runs independently from the calling application. Out-of-process conversions protect the stability of the calling application in the rare case when a malformed document causes Export to fail. [Convert Files Out of Process, on page 24](#).

Subfile Extraction

To convert a file, you must first determine whether the source file contains any subfiles (attachments, embedded objects, and so on). A file that contains subfiles is called a *container* file. Compressed files (such as Zip), mail messages with attachments (such as Microsoft Outlook Express), mail stores (such as Microsoft Outlook Personal Folders), and compound documents with embedded OLE objects (such as a Microsoft Word document with an embedded Excel chart) are examples of container files.

If the file is a container file, the container must be opened and its subfiles extracted by using the *File Extraction API*. The extraction process is done repeatedly until all subfiles are extracted and exposed for conversion. After you extract a subfile, you can use the HTML Export API to convert the file.

If a file is not a container, you should pass it directly to the HTML Export API for conversion without extraction.

See [Use the File Extraction API, on page 47](#) for more information.

Convert Outlook Email without Using the Extraction API

Micro Focus strongly recommends that you convert all container files, including Microsoft Outlook files, by using the File Extraction API. However, you can convert Outlook email messages (MSG) directly by using the Export API and the MSG reader (msgsr).

NOTE:

The MSG reader extracts only the message body of an MSG file. Attachments are not extracted.

To convert MSG files by using the MSG reader, add the following to the `formats_e.ini` file (TRUE is case-sensitive):

```
[ContainerOptions]
bConvertMSG=TRUE
```

Set Conversion Options

Conversion options are parameters that determine the content, structure, and appearance of the HTML output. For example, you can specify

- the markup inserted at the beginning and end of specific HTML blocks
- whether a heading is included in the table of contents
- the output character set
- the resolution at which graphics are converted

You can set the conversion options either in the API or in the template files. Regardless of the method used to set the options, the values are ultimately passed to the API and used to populate the following classes:

- `HtmlHeadingInfo`
- `HtmlOptionInfo`
- `HtmlTemplateInfo`
- `HtmlTOCOptionInfo`

Set Conversion Options by Using the API

Use the following classes to set conversion options:

- `HtmlHeadingInfo`
- `HtmlOptionInfo`
- `HtmlTemplateInfo`
- `HtmlTOCOptionInfo`

There are methods to get or set the corresponding objects in the `HtmlExport` instance. In the `HtmlExport` class, the following methods are available:

- `getOptionInfo/setOptionInfo`
- `getTemplateInfo/setTemplateInfo`
- `getTOCOptionInfo/setTOCOptionInfo`

In the `HtmlTOCOptionInfo` class, the following methods are available:

- `getHtmlHeadingInfo`
- `setHtmlHeadingInfo`

Set Conversion Options by Using the Template Files

HTML Export includes templates in the form of initialization files (`.ini`). The templates provide a quick and easy way to modify the conversion options without programming at the API level. However, the template files do not give you complete control of the conversion process. To control some features, you must use the API directly.

You can use a text editor to fully customize the template files. For example, to change the output character set from the default `KVCS_UNKNOWN` to `KVCS_SJIS` in the `default.ini` template, make the change shown in bold below:

```
[KVHTMLOptionsEx]
OutputCharSet=KVCS_SJIS
bUseDocumentColors=TRUE
```

To create valid HTML, a template file must define at least two sections: `KVHTMLTemplateEx` and `KVHTMLOptionsEx`.

NOTE: If you enter markup in the template files that is not compliant with HTML standards, HTML Export inserts the markup into the output file unchanged. This might result in a malformed HTML file.

An application must then read the template file and write the data to the appropriate Export class. In the `HtmlTest` sample program, a template file is supplied as a command-line argument (see [HtmlTest, on page 112](#)). The template file is passed using the `setIniFileName` method.

```
objHtmlExport.setIniFileName(iniFile);
```

The characteristics of some of the template files are demonstrated in the HTML Export Getting Started page. The Getting Started page, named `htmstart.html`, is in the directory `install\htmlexport\docs`, where `install` is the path name of the Export installation directory. It compares the output generated using a set of sample documents and the template files. The source documents used in the page are in the directory `install\testdocs`.

Templates

The template files for the Java API implementation are in the directory `install\javaapi\ini\html`, where `install` is the path name of the Export installation directory.

The following templates are provided for the Java implementation.

| Template | Description |
|--|---|
| Callback (callback.ini) | <ul style="list-style-type: none">Based on the default template (default.ini).Implements a user callback named "UserCB_End_Block " at the bottom of the main HTML file. |
| Single file with table of contents (onefiletoc.ini) | <p>This template is useful when you want to print the document.</p> <ul style="list-style-type: none">Creates a single HTML file.Creates a table of contents at the top of the HTML document.Uses worksheet names to create the table of contents entries for spreadsheets. If worksheet names do not exist in the source document, "Sheet1," "Sheet2," "Sheet3," and so on are used.Uses slide titles to create the table of contents entries for presentations. If slide titles do not exist in the source document, "slide 1," "slide 2," "slide 3," and so on are used.Lists all metadata (title, subject, author, comments, and so on).Converts graphics to JPEG with the original resolution preserved.Converts presentation slides to HTML as individual JPEG files. |
| UNIX web server (defunix.ini) | <ul style="list-style-type: none">Based on the default template (default.ini).Converts embedded graphics or presentations to either JPEG or HTML Export's Java target. See Display Vector Graphics on UNIX and Linux, on page 83. |

Use the Export Demo Program

The easiest way to get started with Export is to become familiar with its capabilities through the Visual Basic sample program, Export Demo.

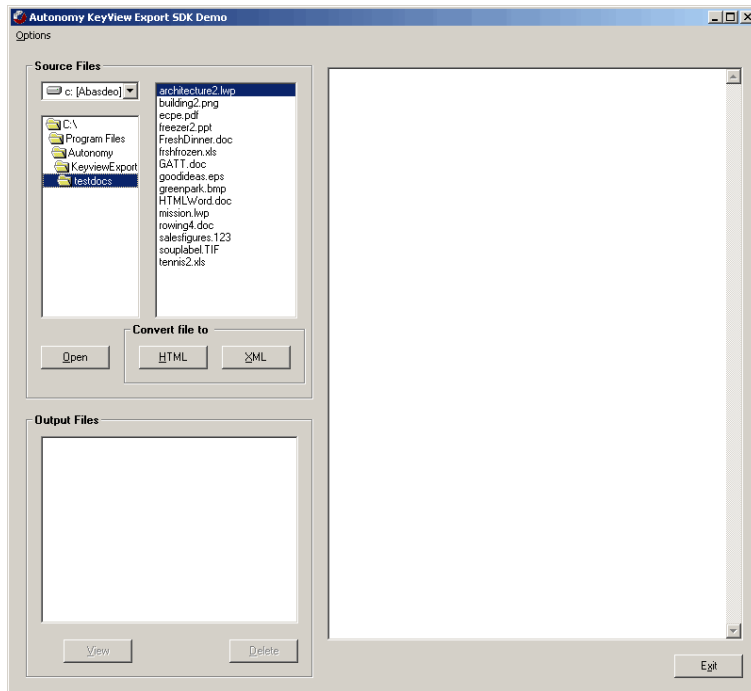
The source code for the program is in the directory *install*\htmlexport\programs\ExportDemo, where *install* is the path name of the Export installation directory.

Export Demo is for Windows only, and requires Internet Explorer 4.01 with Service Pack 1 or higher.

The output options that control the look of the output files are predefined in Export Demo and cannot be changed in the user interface. Export Demo uses a small sample of the options available in the Export API. The Template Wizard sample program is an example of a Visual Basic program that does allow the user to control some of the output options by using template files. You can use the sample documents in the directory *install*\testdocs to experiment with converting different file formats.

To launch the sample program, select **Export Demo** from **Start > Programs > Autonomy > Export SDK > HTML Export**. The following dialog box appears:

Export Demo: Launching



NOTE:

XML conversion using XML Export is available in Export Demo if you have XML Export installed. If you do not have XML Export installed, the **XML** button is unavailable.

Change Input/Output Directories

If HTML Export is installed in the default directory, the output and input directories are automatically set.

The default location for source files is the *install\testdocs* directory.

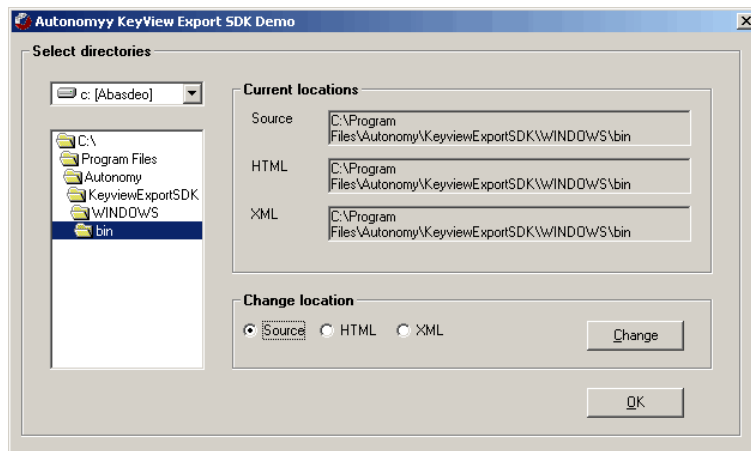
The default location for output files is the *install\htmlexport\programs\tempout* directory.

If HTML Export is installed in a directory other than the default, you are prompted to select an output and input directory when you first start Export Demo.

To change the default directories for the source and output files

1. Select **Options > Set Directories**.

Export Demo: Setting Directories



2. From the tree view, select the drive letter and directory for the source or output files.
3. In **Change Location**, select which files are stored in the directory, either **Source** or **HTML**.
4. Click **Change**. The **Current Locations** fields are updated with the new selection.
5. Follow the same procedure for the other file types. When you are finished, click **OK**.

Set Configuration Options

With HTML Export, you can configure options prior to the document conversion. Export Demo demonstrates this functionality, and enables you to:

- Specify a PDF reader
- Specify whether rotated text in a PDF file is displayed in its original position or at the bottom of the page

Convert PDF Files

In Export Demo, you can convert PDF documents in one of two ways:

- generate HTML output by using the basic PDF reader (`pdfsr`)
- generate a JPEG for each page of the PDF file by using a graphic-based PDF reader (`kppdfldr` or `kppdf2rdr`). See [Convert PDF Files to Raster Images, on page 87](#).

Export Demo provides an option to select the type of reader that you want to use to convert PDF documents. By default, the basic reader (`pdfsr`) is used to convert PDF documents.

To specify that the graphic-based reader be used to convert PDF documents

1. Ensure that Export Demo is not running.
2. Set the appropriate configuration file options. See [Convert PDF Files to Raster Images, on page 87](#).
3. Start the Export Demo program.

4. Select **Options > HTML Config > Set Hifi**.

NOTE:

You can also convert PDF documents to a logical reading order by using the `pdfsr` PDF reader. This feature is demonstrated in the `cnv2html` C sample program, and the `Htm1Test` Java sample program. [Convert PDF Files to a Logical Reading Order, on page 88](#).

Convert Rotated Text

In HTML Export, you can specify how rotated text is displayed in the HTML output. By default, rotated text in a file is displayed in its original position, at the original font size, and at zero degrees rotation. Because the text is the original size, but might be displayed in a smaller space, the text might overlap adjacent text in the HTML output. You use the text rotation configuration option to avoid this problem. If you set this option, rotated text is displayed at the bottom of the page on which it appears. See [Convert Rotated Text, on page 92](#). Currently, this configuration option applies only to PDF files.

To specify that rotated text be displayed at the bottom of the page on which it appears, select **Options > HTML Config > Set Text Rotate**.

Convert Files

To convert a single file

1. Select **Options > Convert > Single file**.
2. Select the document from the file list. Click **HTML** in the **Convert file to** pane.

To convert files in a directory

1. Select **Options > Convert > Entire directory**.
2. Click **HTML** in the **Convert directory to** pane.

To view a converted file, double-click the output file in the **Output Files** pane, or select the output file, and then click **View**. The converted file is displayed in the view pane:

Export Demo: Converting Files



To view the original document, select the document from the file list, and then click **Open**. If you have an application on your system associated with the file, the file is displayed in that application.

To delete output files, select the file in the **Output Files** pane and click **Delete**.

Use the HTML Export API

The Java version of the API provides an interface to the core functionality of the C API. It contains one primary class (`HtmlExport`) that wraps the conversion functionality of the C API. The `HtmlExport` class provides access to a family of methods called `convert`.

The API is implemented in the `com.verity.api.export` package, which is contained in the `KeyView.jar` file. The jar file is in the `install\javaapi` directory, where `install` is the path name of the Export installation directory.

For a full description of the Java API classes, see the Javadoc in the directory `install\javaapi\javadoc`. Sample programs are provided to demonstrate the Java API. See [Sample Programs, on page 109](#).

You can access the Java API directly or by setting conversion options in template files, which are then passed to the API. For more information on template files, see [Set Conversion Options, on page 31](#).

For information on the C API, see the *HTML Export C Programming Guide*.

Input/Output Operations

In the Export Java API, input and output can be either a physical file accessed through a file path, or a Java stream. Depending on the `convert` method signature that you use, you can create the following conversion processes:

- convert an input file to output file
- convert an input file to an output stream
- convert an input stream to an output stream
- convert an input stream to an output file

You can set the input source by calling the `setInputSource` method, or when using the `convert` method. The latter takes the input source as one of its parameters. When you select a `convert` method, ensure that you use the correct signature for the desired input and output type.

NOTE:

When the input source is from a Java stream, Export creates an internal buffer from the stream. If the input is a large file, Micro Focus recommends that you use a file as the input source.

Convert Files

To convert files using the methods in the `HtmlExport` class

1. Instantiate an `HtmlExport` object.

```
m_objExport = new HtmlExport();
```

2. If you are using a template file to set the conversion options, call the `setIniFileName` method. It is recommended you set the full path to the template file.

```
m_objExport.setIniFileName(m_iniFile);
```

Conversion options are parameters that determine the content, structure, and appearance of the HTML output. See [Set Conversion Options, on page 31](#).

3. If you are using the API to set the conversion options, create instances of the following classes:

- `HtmlOptionInfo`
- `HtmlTemplateInfo`
- `HtmlTOCOptionInfo`
- `StyleMapping`
- `HtmlHeadingInfo`

Conversion options are parameters that determine the content, structure, and appearance of the HTML output. [Set Conversion Options, on page 31](#).

Set the classes to the current `HtmlExport` object using the following methods available in the `HtmlExport` class:

- `setOptionInfo`
- `setTemplateInfo`
- `setTOCOptionInfo`
- `setStyleMapping`

and the following method available in the `HtmlTOCOptionInfo` class:

- `setHtmlHeadingInfo`

4. Set the location of the Export libraries by calling the `setExportDirectory` method. These are normally in the directory `install\OS\bin`, where `install` is the path name of the Export installation directory and `OS` is the name of the operating system.

```
m_objExport.setExportDirectory(m_exportDirectory);
```

5. Open the source file by calling the `extOpenDocument` method. This call defines the parameters necessary to open a file for extraction.

```
ExtOpenDocConfig odconfig = null;
long extContextID = 0;
odconfig = new ExtOpenDocConfig();
odconfig.setUserName(m_userName);
odconfig.setPassword(m_password);
odconfig.setUserIDFile(m_userIDFile);
odconfig.setExtractDirectory(m_extractDir);
odconfig.setCreateRootNode(m_createRootNode);
extContextID = m_objExport.extOpenDocument(inFile, odconfig);
```

6. Determine whether the main file is a container file (that is, whether it contains subfiles) by calling the `extGetMainFileInfo()` method.

```
ExtMainFileInfo maininfo = null;
maininfo = m_objExport.extGetMainFileInfo(extContextID);
```

7. If the call to `extGetMainFileInfo()` determined that the source file is a container file, proceed to [Step 8](#); otherwise, proceed to [Step 11](#).

8. Determine whether the subfile is itself a container (that is, whether it contains subfiles) by calling the `extGetSubFileInfo` method.

```
ExtSubFileInfo subinfo = null;
if(nSubFiles != 0)
{
    for(int index = 0; index < nSubFiles; index++)
    {
        subinfo = m_objExport.extGetSubFileInfo(extContextID, index);
        ...
    }
}
```

9. Extract the subfile by calling the `extExtractSubFile` method.

```
ExtSubFileExtractConfig extconfig = null;
extconfig = new ExtSubFileExtractConfig();
extconfig.setCreateDirectory(true);
extconfig.setOverWrite(true); extconfig.setExcludeMailHeader(m_
excludeMailHeader);
extinfo = m_objExport.extExtractSubFile(extContextID, index, extconfig);
```

10. If the call to `extGetSubFileInfo` determined that the subfile is a container file, repeat [Step 5](#) through [Step 9](#) until all subfiles are extracted; otherwise, proceed to [Step 11](#).
11. Optionally, set the input source as either a file or input stream by calling the `setInputSource` method.

```
if(m_inputMethod == Export.IO_METHOD_FILE)
{
    //input as file
    m_objExport.setInputSource(m_extractDir + filename);
}
else
{
    //input as stream
    File inf = new File(m_extractDir + filename);
    FileInputStream fis = new FileInputStream(inf);
    m_objExport.setInputSource(fis);
    fis.close();
}
```

12. Set up an out-of-process session by calling the `startOOPSession` method. This initializes the out-of-process session, creates a Servant process, establishes a communication channel between the application thread and the Servant, and sends the data to the Servant.

```
m_objExport.startOOPSession();
```

13. Convert the input and generate the output files by calling the `convertTo` method. You cannot use the `convert` methods that set the input source because the input source must be set *before* the out-of-process session is initialized. The `convertTo` method can only be called once in a single out-of-process session.

```
if(m_outputMethod == Export.IO_METHOD_FILE)
{
    //convert to a file
    m_objExport.convertTo(m_extractDir + filename + m_extension);
}
else
{
    //convert to a stream
    File outf = new File(m_extractDir + filename + m_extension);
    FileOutputStream fos = new FileOutputStream(outf);
    m_objExport.convertTo(fos);
}
```

```
        fos.close();  
    }
```

14. If you are converting additional files, terminate the out-of-process session by calling the `endOOPSession` method and setting the Boolean to `TRUE`. The Servant ends the current conversion session and releases the source data and session resources.

If you are not converting additional files, terminate the out-of-process session and the Servant process by calling `endOOPSession` and setting the Boolean to `FALSE`.

```
if(i == (nSubFiles - 1))  
{  
    m_keepServantAlive = false;  
}  
else  
{  
    m_keepServantAlive = true;  
}  
m_objExport.endOOPSession(m_keepServantAlive);
```

15. Close the file by calling the `extCloseDocument()` method.

```
m_objExport.extCloseDocument(extContextID);
```

16. Repeat [Step 2](#) through [Step 15](#) for additional source files.

17. Terminate the session and free allocated system resources by calling the `shutdownExport()` method.

```
m_objExport.shutdownExport();
```

Multithreaded Conversions

To ensure that multithreaded conversions are thread-safe, you must create a unique Export context for every thread by instantiating an `HtmlExport` object. In addition, threads must not share context objects, and you must use the same context object for all API calls in the same thread. Creating a context object for every thread does not affect performance because the context object uses minimal resources.

For example, your Java code should have the following logic in a thread:

```
m_objExport = new HtmlExport();  
m_objExport.setIniFileName(m_iniFile);  
m_objExport.setExportDirectory(m_exportDir);  
m_objExport.extOpenDocument(inFile, odconfig);  
m_objExport.extGetMainFileInfo(extContextID) /* container file */  
m_objExport.extGetSubFileInfo(extContextID, index);  
m_objExport.extExtractSubFile(extContextID, index, extconfig);  
m_objExport.startOOPSession();
```

```
m_objExport.convertTo(outFile);  
m_objExport.endOOPSession(bKeepServantAlive TRUE);  
m_objExport.extCloseDocument();  
m_objExport.extOpenDocument(inFile, odconfig);  
m_objExport.extGetMainFileInfo(extContextID); /* not a container file */  
m_objExport.startOOPSession();  
m_objExport.convertTo(outFile);  
m_objExport.endOOPSession(bKeepServantAlive TRUE);  
m_objExport.extCloseDocument();  
...  
m_objExport.shutdownExport();
```

Use Callbacks

Some Export methods enable you to specify a callback, which is called to control the conversion while it is in progress. For example, you can specify a callback to report progress during the conversion.

To use callbacks

1. If you are using the `UserCBCallback` interface, include the `$USERCB` token in a member of `KVHTMLTemplateEx`. For example, placing "`$USERCB=my_callback` " in `pszFirstH1Start` results in a callback at the point when `pszFirstH1Start` is processed. The user callback function is identified by the text assigned to `$USERCB`, which in this example is `my_callback`.

NOTE: The callback identifier must be delimited by a trailing white space. For example, "`my_callback` "

The `callback.ini` template file provides an example of how to use callback tokens. The file is in the `install\javaapi\ini` directory. See [Export Tokens, on page 229](#) for more information on tokens.

2. Implement the callback interfaces. The callback interfaces are:
 - `CallbackConstants`
 - `CallingContext`
 - `ContinueCallback`
 - `GetAnchorCallback`
 - `GetAuxOutputCallback`
 - `UserCBCallback`

Sample implementations of the callback interfaces are in the `install\javaapi\sample\com\verity\api\htmllexport` directory, where *install* is the path name of the Export installation directory.

3. Declare the objects of the callback procedures you are going to use, and pass them to your instance of `HtmlExport`.

Example

```
ContinueCallback cci = new ContinueCallbackImpl();
GetAnchorCallback gaci = new GetAnchorCallbackImpl();
GetAuxOutputCallback gaoci = new GetAuxOutputCallbackImpl();
UserCBCallback ucbi = new UserCBCallbackImpl();
CallingContext cContext = new CallingContextImpl();
objHtmlExport.setCallingContext(cContext);
objHtmlExport.setContinueCallback(cci);
objHtmlExport.setGetAnchorCallback(gaci);
objHtmlExport.setGetAuxOutputCallback(gaoci);
objHtmlExport.setUserCBCallback(ucbi);
```

Before Running Your Application

Before running your application, set the library path using one of the following methods:

- On Windows, add the location of `htmllexport.dll` to the `PATH` environment variable.
- On Solaris, Linux, and HP-UX IA-64, add the location of `libhtmllexport.so` to the `LD_LIBRARY_PATH` environment variable.
- On HP-UX PA-RISC, add the location of `libhtmllexport.sl` to the `SHLIB_PATH` environment variable.
- On AIX, add the location of `libhtmllexport.a` to the `LIBPATH` environment variable. You can also specify the library path as a system property as follows:

```
java -Djava.library.path=HTML_Export_bin_directory ...
```

Export .NET Support

KeyView provides interfaces to its Java APIs in the .NET environment. These interfaces enable you to write C# or Visual Basic .NET programs that call Export APIs and run them natively in the Microsoft .NET Common Language Runtime (CLR).

For a full description of the API classes, see the Javadoc in the directory `install\javaapi\javadoc`.

Third-Party Software and Redistributable Files

You need the following components to develop applications in the .NET environment:

- Microsoft .NET Framework Version 2.0 SDK
- Microsoft .NET Framework Version 2.0 Redistributable Package
- Micro Focus-supplied dynamic link library, `kvexportdotnet`, which is installed to the `Export \bin` directory.

NOTE:

The KeyView installation does not register the Export .NET library in the Global Assembly Cache (GAC) on your machine.

The following files must be redistributed with your .NET application:

- `kvexportdotnet` library (.NET API)
- `kvhtmlwrapper` library (C wrapper around `kvhtml`)
- Microsoft .NET Framework Version 2.0 Redistributable Package
- other files listed in the [Files Required for Redistribution, on page 240](#).

Sample Programs

The following executables are J# sample programs that have been compiled for .NET. They are installed to the `Install_Dir\Export SDK\bin` directory.

- `ExtractExport.exe` (See [ExtractExport, on page 109](#))
- `HtmlTestJ.exe` (See [HtmlTest, on page 112](#))

The following C# program is also available. It is installed to the `Install_Dir\Export SDK\dotnet\sample` directory and can be compiled in Visual Studio:

- `HtmlConvFileToFile.cs` (See [HtmlConvFileToFile, on page 114](#))

Part II: Use the Export API

This section explains how to perform some basic tasks using the File Extraction and Export APIs, and describes the sample programs. It contains the following chapters:

- [Use the File Extraction API](#)
- [Use the HTML Export API](#)
- [Sample Programs](#)

Chapter 3: Use the File Extraction API

This section describes how to extract subfiles from a container file using the File Extraction API.

- [Introduction](#)47
- [Extract Subfiles](#)48
- [Extract Images](#)50
- [Recreate a File Hierarchy](#)50
- [Extract Mail Metadata](#)52
- [Extract Subfiles from Outlook Files](#)57
- [Extract Subfiles from Outlook Express Files](#)57
- [Extract Subfiles from Mailbox Files](#)57
- [Extract Subfiles from Outlook Personal Folders Files](#)58
- [Extract Subfiles from Lotus Domino XML Language Files](#)61
- [Extract Subfiles from Lotus Notes Database Files](#)62
- [Extract Subfiles from PDF Files](#)65
- [Extract Embedded OLE Objects](#)65
- [Extract Subfiles from ZIP Files](#)66
- [Default File Names for Extracted Subfiles](#)66
- [Exclude Japanese Guide Text](#)68

Introduction

To convert a file, you must first determine whether the file contains any subfiles (attachments, embedded OLE objects, and so on). A file that contains subfiles is called a *container* file. A container file has a main file (parent) and subfiles (children) embedded in the main file.

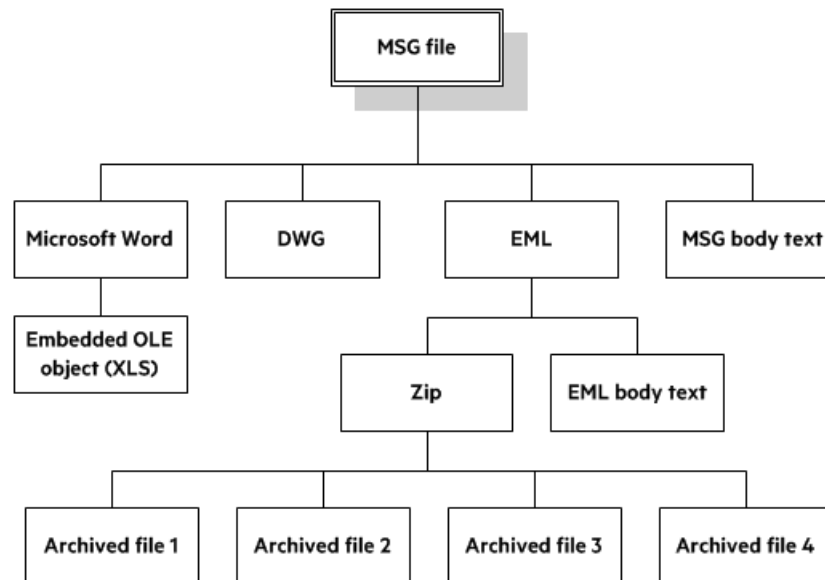
The following are examples of container files:

- Archive files such as ZIP, TAR, and RAR.
- Mail messages such as Outlook (MSG) and Outlook Express (EML).
- Mail stores such as Microsoft Outlook Personal Folders (PST), Mailbox (MBX), and Lotus Notes database (NSF).
- PDF files that contain file attachments.
- Compound documents with embedded OLE objects such as a Microsoft Word document with an embedded Excel chart.

NOTE: [Supported Formats](#), on page 121 indicates which formats are treated as container files and are supported by the File Extraction API.

The subfiles might also be container files, creating a file hierarchy of multiple levels. For example, an MSG file (the root parent) might contain three attachments:

- a Microsoft Word document that contains an embedded Microsoft Excel spreadsheet.
- an AutoCAD drawing file (DWG).
- an EML file with an attached Zip file, which in turn contains four archived files.



NOTE: The parent MSG file contains four first-level children. The body text of a message file, although not a standalone file in the container, is considered a child of the parent file.

Extract Subfiles

To convert all files in a container file, the container must be opened and its subfiles extracted to either a file or a stream using the *File Extraction API*. The extraction process is done repeatedly until all subfiles are extracted and exposed for conversion. Once a subfile is extracted, you can call Export API methods to convert the data.

If you require a container file, including subfiles, to be converted to a single file, you must extract all files from the container, convert the files, and then append each converted output to its parent.

To extract subfiles, follow this general procedure

1. Open the source file by calling the `extOpenDocument` method. This call defines the parameters necessary to open a file for extraction.
2. Determine whether the main file is a container file (contains subfiles) by calling the `extGetMainFileInfo()` method.
3. If the call to `extGetMainFileInfo()` determined the source file is a container file, proceed to [Step 4](#); otherwise, convert the file.

4. Determine whether the subfile is itself a container (contains subfiles) by calling the `extGetSubFileInfo` method.
5. Extract the subfile by calling the `extExtractSubFile` method.
6. If the call to `extGetSubFileInfo` determined the subfile is a container file, repeat [Step 1](#) through [Step 5](#) until all subfiles are extracted and the lowest level of subfiles is reached; otherwise, convert the file.

Sanitize Absolute Paths

When you extract a subfile from a container and write it to disk, you specify an extract directory and a path to extract the file to.

To set the path, you might use the path in the container file that you are extracting from, as returned from the `Filter.extGetSubFileInfo()` method. However, if the path is an absolute path, the file could be created outside the directory you have chosen as the extract directory. Your application might then contain a vulnerability that could be exploited to write files to unexpected locations in the file system. This section discusses some KeyView features that can help you secure your application by sanitizing paths.

KeyView always sanitizes relative paths that you pass in when extracting files, so that the paths remain within the extract directory you specify. For example, KeyView does not allow the use of `..` to move outside the extract directory.

KeyView can update absolute paths so that they remain within the extract directory. You can instruct KeyView to sanitize absolute paths programmatically (through the API), or by setting a parameter in the configuration file.

The following table shows the effect on some example paths.

| Requested path | Path of extracted file (not sanitized) | Path of extracted file (sanitized) |
|----------------|--|------------------------------------|
| file.txt | <i>extractDir/file.txt</i> | <i>extractDir/file.txt</i> |
| dir/file.txt | <i>extractDir/dir/file.txt</i> | <i>extractDir/dir/file.txt</i> |
| ../file.txt | <i>extractDir/file.txt</i> | <i>extractDir/file.txt</i> |
| /dir/file.txt | <i>/dir/file.txt</i> | <i>extractDir/dir/file.txt</i> |

To sanitize absolute paths

- Call the method `setSanitizeAbsolutePath` on the `ExtSubFileExtractConfig` that you pass in to `extExtractSubFile`. When KeyView sanitizes a path and the resulting directory does not exist, extraction fails unless you instruct KeyView to create the directory, so you might also want to call the method `setCreateDirectory`. You can find the path that a file was actually extracted to from the `ExtSubFileExtractInfo` object that is returned from the `extExtractSubFile` method.

To sanitize absolute paths (through configuration)

- In the `formats_e.ini` configuration file, set the parameter `SanitizeAbsoluteExtractPaths`, for example:

```
[Options]
SanitizeAbsoluteExtractPaths=TRUE
```

Extract Images

You can use the File Extraction API to extract images within the file by specifying the following in the `formats.ini` file:

```
[Options]
ExtractImages=TRUE
```

If you set this option, images within the file behave in the same way as any other subfile. Extracted images have the name `image[X].[Y]`, where `[X]` is an integer, and `[Y]` is the extension. The format of the image is the same as the format in which it is stored in the document.

This option can also be enabled by passing `KVFLT_EXTRACTIMAGES` to the `fpFilterConfig` function.

NOTE:

Turning on `ExtractImages` can reduce the speed of the filtering operation.

Recreate a File Hierarchy

When a container file is extracted, any relationships between the subfiles in the container are not maintained. However, the File Extraction interface provides information that enables you to recreate the hierarchy. The hierarchy can be used to create a directory structure in a file system, or to categorize documents according to their relationship to each other. For example, if you use `KeyView` to generate text for a search engine, the hierarchical information enables your users to search for a document based on the document's parent or sibling. In addition, when the document is returned to the user, the parent and sibling documents can be returned as recommendations.

The information needed to recreate a file's hierarchy is provided in the call to `extGetSubFileInfo`. Call this method to retrieve an object of the `ExtSubFileInfo` class, then use the `getParentIndex()` and `getChildArray()` methods in this object to retrieve information about the subfile's parent and children. Since you can only retrieve the first-level children in a subfile, you must call `extGetSubFileInfo` repeatedly until information for the leaf-node children is extracted.

Create a Root Node

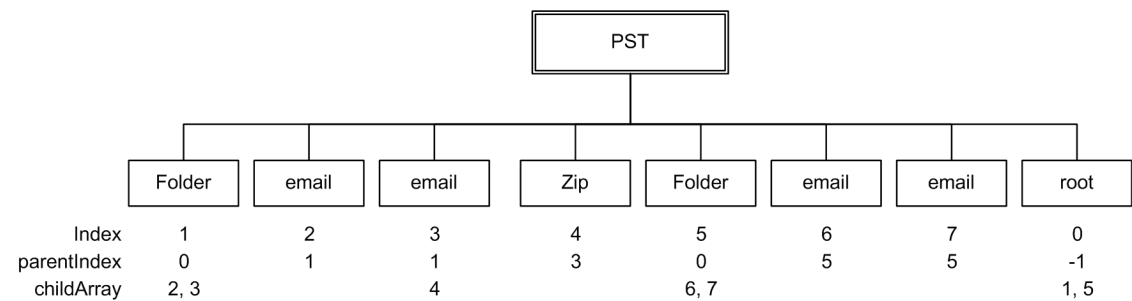
Because of their structure, some container files do not contain a subfile or folder which acts as a root directory on which the hierarchy can be based. For example, subfiles in a Zip archive can be extracted, but none of the subfiles represent the root of the hierarchy. In this case, an artificial *root node* must be created at the top of the file hierarchy as a point of reference for each child, and ultimately to recreate the relationships. This artificial root node is an internal object, and is extracted to disk as a directory called `root`. Its index number is 0.

To create a root node, call the `setCreateNode` method in the `ExtOpenDocConfig` object, and pass `ExtOpenDocConfig` to the `extOpenDocument` method. When a root node is created, the value returned from the `getNumSubFiles` method in the `ExtMainFileInfo` object includes the root node. For example, when you call `extGetMainFileInfo` on a Microsoft Word document with three embedded OLE objects and the root node is disabled, the number of subfiles is 3. If you create a root node, the number of subfiles is 4.

Example

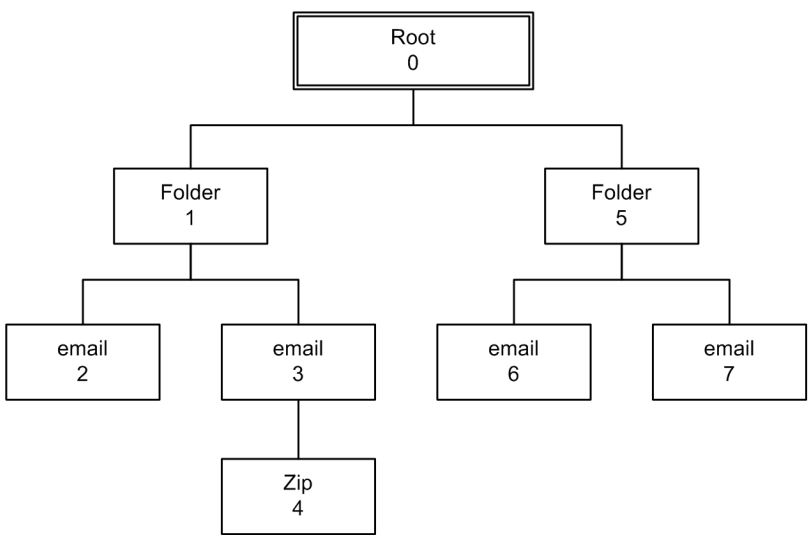
For example, you might extract a PST file that contains seven subfiles with a root node enabled. The call to `extGetMainFileInfo()` returns the number of subfiles as 8 (seven subfiles and one root node). The following diagram shows the structure and the available hierarchy information after the subfiles are extracted:

Extracted PST File



The `parentIndex` specifies the index number of a subfile's parent. The `childArray` specifies an array of a subfile's children. With this information, you can recreate the hierarchy shown in the following diagram:

Recreated File Hierarchy



Extract Mail Metadata

You can extract metadata such as subject, sender, and recipient from subfiles of mail formats by calling the `extGetSubFileMetadata()` method. You can extract a predefined set of metadata fields, or a list of metadata fields by their names or MAPI properties.

Default Metadata Set

KeyView internally defines a set of common mail metadata fields that can be extracted as a group from mail formats. This default metadata set is listed in the following table.

Default Mail Metadata List

| Field Name (string to specify) | Description |
|--------------------------------|---|
| From | The display name and email address of the sender. |
| Sent | The time the message was sent. |
| To | The display names and email addresses of the recipients. |
| Cc | The display names and email addresses of recipients who receive copies of the email. |
| Bcc | The display names and email addresses of recipients who received blind copies of the email. |
| Subject | The text in the subject line of the message. |
| Priority | The priority applied to the message. |

Because mail formats use different terms for the same fields, the format's reader maps the default field name to the appropriate format-specific name. For example, when retrieving the default metadata set, the NSF field *Importance* is mapped to the name *Priority* and is returned.

You can also extract the default field names individually by passing the field name (such as *From*, *To*, and *Subject*); however, in this case, the string is not mapped to the format-specific name. For example, if you pass *Priority* in the call, you will retrieve the contents of the *Priority* field from an MBX file, but will not retrieve the contents of the *Importance* field from an NSF file.

NOTE: You cannot pass the field names listed in [MSG-Specific Metadata List, on the next page](#) individually for PST files. However, you can pass either the MAPI tag number or one of the constants in the Export class as integers. See [Microsoft Personal Folders File \(PST\) Metadata, on page 55](#).

Extract the Default Metadata Set

To extract the default metadata set, call the `extGetSubFileMetadata(long docContextID, int nSubFileIndex, ExtSubFileMetaConfig config)` method.

For example:

```
ExtSubFileMetaConfig metaConfig = new ExtSubFileMetaConfig();  
ExtSubFileMetadata subfilemeta = null;  
subfilemeta = m_objExport.extGetSubFileMetadata(extContextID, index, metaConfig);
```

Microsoft Outlook (MSG) Metadata

In addition to the default metadata set, the metadata fields listed in the following table can be extracted for MSG files. The field name must be passed to `metaNameArray` in the call to the `extGetSubFileMetadata()` method.

MSG-Specific Metadata List

| Field Name (string to specify) | Description |
|--------------------------------|--|
| AttachFileName | An attachment's long file name and extension, excluding path. |
| ConversationTopic | The topic of the first message in a conversation thread. A conversation thread is a series of messages and replies. This is the first message's subject with any prefix removed. |
| CreationTime | The time the message or attachment was created. This value is displayed in the Sent field in the message's Properties dialog in Outlook. |
| InternetMessageID | The identifier for messages that come in over the Internet. This is the MAPI property <code>PR_INTERNET_MESSAGE_ID</code> . This property is not in the MAPI headers or MAPI documentation. |
| LastModificationTime | The time the message or attachment was last modified. This value is displayed in the Modified field in the message's Properties dialog in Outlook. |
| Location | The physical location of the event specified in the Outlook calendar entry. |
| MessageID | The message transfer system (MTS) identifier for the message transfer agent (MTA). This value is displayed on the Message ID tab in the message's Properties dialog in Outlook. |
| Received | The date and time a message was delivered. This value is displayed in the Received field in the message's Properties dialog in Outlook. |
| Sender | <p>The name and email address of the message sender. This value is a concatenation of two MAPI properties in the following format:</p> <p>"PR_SENDER_NAME" <PR_SENDER_EMAIL_ADDRESS></p> <p>The Sender value might be the same as or different than the default metadata <code>From</code> value (see Default Metadata Set, on the previous page), depending on which MAPI properties exist in the MSG file.</p> |
| Sensitivity | The value indicating the message sender's opinion of the sensitivity of a |

MSG-Specific Metadata List, continued

| Field Name (string to specify) | Description |
|--------------------------------|--|
| | message, such as Personal, Private, or Confidential. This value is displayed in the Sensitivity field in the message's Properties dialog in Outlook. |
| TransportMsgHeaders | Contains transport-specific message envelope information. This value corresponds to the MAPI property PR_TRANSPORT_MESSAGE_HEADERS. |
| StartDate | Contains an appointment start date. This value corresponds to the PR_START_DATE MAPI property. |
| EndDate | Contains an appointment end date. This value corresponds to the PR_END_DATE MAPI property. |

Extract MSG-Specific Metadata

To extract specific metadata fields from an MSG file, use the method `extGetSubFileMetadata(long docContextID, int nSubFileIndex, java.lang.String[] metaNameArray, ExtSubFileMetaConfig config)` and pass the field name defined in [MSG-Specific Metadata List, on the previous page](#) to `metaNameArray` (the string is not case sensitive).

For example, the following code extracts the contents of the `ConversationTopic` and `MessageID` fields:

```
ExtSubFileMetaConfig metaConfig = new ExtSubFileMetaConfig();  
  
ExtSubFileMetadata subfilemeta = null;  
  
String[] metaNameArray = {"conversationtopic", "MessageID"};  
  
subfilemeta = m_objExport.extGetSubFileMetadata(extContextID, index, metaNameArray,  
metaConfig);
```

Microsoft Outlook Express (EML) and Mailbox (MBX) Metadata

In addition to the default metadata set, you can extract any metadata field that exists in the header of an EML or MBX file by passing the field's name. If the name is a valid field in the file, the contents of the field are returned. For example, to retrieve the name of the last mail server that received the message before it was delivered, you can pass the string "Received".

Extract EML- or MBX-Specific Metadata

To extract specific metadata fields from an EML or MBX file, use the method `extGetSubFileMetadata(long docContextID, int nSubFileIndex, java.lang.String[] metaNameArray, ExtSubFileMetaConfig config)` and pass the metadata name to `metaNameArray` (the string is not case sensitive).

For example, the following code extracts the contents of the `Received` and `Mime-version` fields:

```
ExtSubFileMetaConfig metaConfig = new ExtSubFileMetaConfig();  
ExtSubFileMetadata subfilemeta = null;  
String[] metaNameArray = {"Received", "Mime-version"};  
subfilemeta = m_objExport.extGetSubFileMetadata(extContextID, index, metaNameArray,  
metaConfig);
```

Lotus Notes Database (NSF) Metadata

In addition to the default metadata set, you can extract any Lotus field name that exists in an NSF file by passing the field's name. (You can extract fields from mail NSF files and non-mail NSF files.) If the name is a valid field in the file, the field is returned. For example, to retrieve the date a document in an NSF file was last accessed, you would pass the string "\$LastAccessedDB".

NOTE: A complete list of NSF fields are provided in the Lotus Notes file `stdnames.h`. This header file is available in the Lotus API Toolkit.

Extract NSF-Specific Metadata

To extract specific metadata fields from an NSF file, use the method `extGetSubFileMetadata(long docContextID, int nSubFileIndex, java.lang.String[] metaNameArray, ExtSubFileMetaConfig config)` and pass the metadata name to `metaNameArray` (the string is not case sensitive).

For example, the following code extracts the contents of the Description and Categories fields:

```
ExtSubFileMetaConfig metaConfig = new ExtSubFileMetaConfig();  
ExtSubFileMetadata subfilemeta = null;  
String[] metaNameArray = {"description", "Categories"};  
subfilemeta = m_objExport.extGetSubFileMetadata(extContextID, index, metaNameArray,  
metaConfig);
```

Microsoft Personal Folders File (PST) Metadata

In addition to the default metadata set, you can extract Messaging Application Programming Interface (MAPI) properties from a PST file. These properties describe elements (subject, sender, recipient, and so on) of Outlook items within the PST file. Since the properties are stored in the PST file itself, they can be retrieved before the contents of the PST are extracted. This enables you to determine whether an Outlook item should be extracted based on a subfile's attributes. MAPI properties are also stored for Outlook attachments that are not mail messages (such as an attached Microsoft Word document or Lotus 1-2-3 file).

MAPI Properties

Each MAPI property is identified by a property tag, which is a constant that contains the property type and a unique identifier. For example, the property that indicates whether a message has attachments

has the following components:

| | |
|---------------|-------------------|
| Property | PR_HASATTACH |
| Identifier | 0x0E1B |
| Property type | PT_BOOLEAN (000B) |
| Property tag | 0x0E1B000B |

The Microsoft MAPI documentation on the Microsoft Developer Network website lists all available MAPI properties, their tags, and types.

You can retrieve any MAPI property that is of one of the MAPI property types listed below:

| | | |
|------------|-----------|------------|
| PT_I2 | PT_DOUBLE | PT_STRING8 |
| PT_I4 | PT_FLOAT | PT_TSTRING |
| PT_BINARY | PT_LONG | PT_SYSTIME |
| PT_BOOLEAN | PT_SHORT | PT_UNICODE |

NOTE: Properties with a PT_TSTRING type have the property type recompiled to either a Unicode string (PT_UNICODE) or to an ANSI string (PT_STRING8) depending on the operating system's character set. To retrieve the Unicode property, pass in the Unicode version of the tag. For example, the property tag for PR_SUBJECT is either 0x0037001E for an ANSI string, or 0x0037001F for a Unicode string.

Extract PST-Specific Metadata

In the call to extract subfile metadata, you can pass either the MAPI tag number (such as 0x0070001e) or one of the constants in the Export class (such as KVPR_SUBJECT). These constants are a subset of MAPI properties and use a KeyView naming convention. For example, the property PR_CONVERSATION_TOPIC is defined as KVPR_CONVERSATION_TOPIC. If the property you want to retrieve is not defined as a constant in the Export class, you must pass the MAPI tag number.

To extract specific MAPI properties from a PST file, use the method `extGetSubFileMetadata(long docContextID, int nSubFileIndex, int[] metaNameArray, ExtSubFileMetaConfig config)` and pass the tag number or constant to `metaNameArray`.

For example, the following code extracts the MAPI properties PR_SUBJECT and PR_ALTERNATE_RECIPIENT:

```
ExtSubFileMetaConfig metaConfig = new ExtSubFileMetaConfig();

ExtSubFileMetadata subfilemeta = null;

int[] metaNameArray = {Export.KVPR_SUBJECT, 0x3A010102};

subfilemeta = m_objExport.extGetSubFileMetadata(extContextID, index, metaNameArray,
metaConfig);
```

Exclude Metadata from the Extracted Text File

When a mail message is extracted, the message text and header information (To, From, Sent, and so on) is also extracted. You can prevent the header information from appearing in the text file.

To exclude the header information, call the `setExcludeMailHeader()` method of the `ExtSubFileExtractConfig` object, and pass `ExtSubFileExtractConfig` to the `extExtractSubFile` method. For example:

```
m_excludeMailHeader = true;

extconfig = new ExtSubFileExtractConfig();

extconfig.setExcludeMailHeader(m_excludeMailHeader);

extinfo = m_objExport.extExtractSubFile(extContextID, i, extconfig);
```

Extract Subfiles from Outlook Files

When you extract an Outlook file (MSG) to disk, the message text and header information (To, From, Sent, and so on) is extracted to a text file. (If you do not want the header information to appear in the text file, see [Exclude Metadata from the Extracted Text File, above](#).) If the Outlook file contains a non-mail attachment, the attachment is extracted in its native format to a subdirectory. If the Outlook file contains a mail attachment, the attachment's message text and any attachments are extracted to a subdirectory.

Extract Subfiles from Outlook Express Files

When you extract an Outlook Express (EML) file to disk, the message text and header information (To, From, Sent, and so on) is extracted to a text file. (If you do not want the header information to appear in the text file, see [Exclude Metadata from the Extracted Text File, above](#).) If the Outlook Express file contains a non-mail attachment, the attachment is extracted in its native format to the same directory as the message text file. If the Outlook Express file contains a mail attachment, the complete attachment (including message text and attachments), the message text file, and any non-mail attachments are extracted to the same directory as the main message.

NOTE: When the MBX reader (`mbxsr`) is enabled, it is used to filter MBX and EML files. If the MBX reader is not enabled, the EML reader (`emlsr`) is used.

Extract Subfiles from Mailbox Files

A Mailbox (MBX) file is a collection of individual emails compiled with RFC 822 and RFC 2045 - 2049 (MIME), and divided by message separators. There are many mail applications that export to an MBX format, such as Eudora Email and Mozilla Thunderbird.

When an MBX file is extracted to disk, the message text and header information (To, From, Sent, and so on) from each mail file are extracted to text files. (If you do not want the header information to appear in the text file, see [Exclude Metadata from the Extracted Text File, above](#).)

In Eudora MBX files, attachments are inserted as a link and are stored externally from the message. These attachments are not extracted, but the path to the attachment is returned in the call to the `extGetSubFileInfo` method. You can write code to retrieve the attachment based on the returned path.

For MBX files from other clients, KeyView extracts attachments when they are embedded in the message.

NOTE: The Mailbox (MBX) reader is an advanced feature and is sold and licensed separately. To enable this reader in a KeyView SDK, you must obtain the appropriate license key from Micro Focus.

Extract Subfiles from Outlook Personal Folders Files

KeyView can extract Outlook items such as messages, appointments, contacts, tasks, notes, and journal entries from a PST file. When a PST file is extracted to disk, the body text and header information (To, From, Sent, and so on) from each Outlook item is extracted to a text file. (If you do not want the header information to appear in the text file, see [Exclude Metadata from the Extracted Text File, on the previous page.](#))

You can also extract messages from PST files as MSG files, including all their attachments, using the `setSaveAsMSG()` method in the `ExtSubFileExtractConfig` class.

If an Outlook item contains a non-mail attachment, the attachment is extracted in its native format to a subdirectory. If an Outlook item contains an Outlook attachment, the attached item's body text and any attachments are extracted to a subdirectory.

NOTE: The Microsoft Outlook Personal Folders (PST) readers are an advanced feature and are sold and licensed separately. To enable these readers in a KeyView SDK, you must obtain an appropriate license key from Micro Focus. For information about adding a new license key to an existing installation, see [Update License Information, on page 18.](#)

Choose the Reader to use for PST Files

KeyView provides several ways of processing PST files:

- Indirectly, using the Microsoft Messaging Application Programming Interface (MAPI). MAPI is a Microsoft interface that enables different applications to exchange messages and attachments with each other. MAPI allows KeyView to open a PST file, traverse the folders, and extract items. The `pstsr` reader uses MAPI, but works only on Windows and requires that Microsoft Outlook is installed.
- Directly, without relying on the Microsoft interface to the PST format. Accessing the file directly does not require Microsoft Outlook. The `pstxsr` reader is available for Windows (32-bit and 64-bit) and Linux (64-bit only). The `pstnsr` reader is an alternative native reader, for the platforms not supported by `pstxsr`.

On Windows, the MAPI-based reader is used by default but you can choose `pstxsr` if you prefer. On UNIX platforms, only one of the native readers is available (`pstxsr` on Linux x64 and `pstnsr` on other platforms).

The differences between the readers are summarized in the following table.

| Feature | Native Reader (pstxsr) | Native Reader (pstnsr) | MAPI-based Reader (pstsrr) |
|-----------------------------------|---|---------------------------------------|---|
| Platforms supported | Windows x86 and x64 Linux x64 | All platforms not supported by pstxsr | Windows x86 and x64 |
| Outlook required | No | No | Yes |
| MAPI properties supported | Yes. All properties defined in <code>mapitags.h</code> . Object properties are not supported. | | |
| Password protection supported | Yes | Yes | Yes (using <code>KVCredential</code> structure) |
| Compressible encryption supported | Yes | Yes | Yes |
| High encryption supported | No | No | Yes |

To change the reader used to process PST files, change the PST entry (file category value 297) in the `formats_e.ini` file. For example, to use `pstxsr`:

```
297=pstx
```

NOTE: You must make sure that the PST that you are extracting is not open in the Outlook client, and that the Outlook process is not running.

NOTE: When extracting subfiles from PST files, information on the distribution list used in an email is extracted to a file called `emailname.dist`. This applies to the MAPI reader (`pstsrr`) only.

System Requirements

MAPI is supported on Windows platforms only and relies on functionality in Outlook. If you want to use the MAPI-based reader, `pstsrr`, Microsoft Outlook must be installed on the same machine as your application. Outlook must also be the default email application. KeyView supports the following PST formats and Outlook clients:

- Outlook 97 or later PST files

NOTE: The Outlook client must be the same version as, or newer than, the version of Outlook that generated the PST file.

- Outlook 2002 or later clients

NOTE: You must install an edition of Microsoft Outlook (32-bit or 64-bit) that matches the KeyView software. For example, if you use 32-bit KeyView, install 32-bit Outlook. If you use 64-bit

KeyView, install 64-bit Outlook.

If the editions do not match, KeyView returns Error 32: KVErrror_PSTAccessFailed and an error message from Microsoft Office Outlook is displayed: Either there is a no default mail client or the current mail client cannot fulfill the messaging request. Please run Microsoft Outlook and set it as the default mail client.

MAPI Attachment Methods

The way in which you can access the contents of a PST message attachment is determined by the MAPI *attachment method* applied to the attachment. For example, if the attachment is an embedded OLE object, it uses the ATTACH_OLE attachment method. KeyView can access message attachments that use the following attachment methods:

ATTACH_BY_VALUE

ATTACH_EMBEDDED_MSG

ATTACH_OLE

ATTACH_BY_REFERENCE

ATTACH_BY_REF_ONLY

ATTACH_BY_REF_RESOLVE

Attachments using the ATTACH_BY_VALUE, ATTACH_EMBEDDED_MSG, or ATTACH_OLE attachment methods are extracted automatically when the PST file is extracted. An "attach by reference" method means that the attachment is not in Outlook, but Outlook contains an absolute path to the attachment. Before you can extract these types of attachments, you must retrieve the path to access the attachment.

To extract "attach by reference" attachments

1. Determine whether the attachment uses an ATTACH_BY_REFERENCE, ATTACH_BY_REF_ONLY, or ATTACH_BY_REF_RESOLVE method by retrieving the MAPI property PR_ATTACH_METHOD.
2. If the attachment uses one of the "attach by reference" methods, get the fully qualified path to the attachment by retrieving the MAPI properties PR_ATTACH_LONG_PATHNAME or PR_ATTACH_PATHNAME.
3. You can then either copy the files from their original location to the path where the PST file is extracted, or use the Export API methods to convert the attachment.

Open Secured PST Files

KeyView enables you to specify credentials (user name and password), which are used to open a secured PST file for extraction.

Detect PST Files While the Outlook Client is Running

If you are running an Outlook client while running the File Extraction API, the KeyView format detection module (*kwad*) might not be able to open the PST file to determine the file's format because Outlook has the file locked. In this case, you can do one of the following:

- Close Outlook when using the Extraction API
- Detect PST files by extension only and bypass the format detection module. To enable this option, add the following lines to the `formats_e.ini` file.

```
[container_flags]
detectPSTbyExtension=1
```

NOTE: The `detectPSTbyExtension` option only applies when you are using the MAPI reader (*pstsr*).

NOTE: If you use this option, you must make sure in your code that valid PST files are passed to KeyView because the format detection module will not be available to verify the file type and pass the file to the appropriate reader.

Extract Subfiles from Lotus Domino XML Language Files

When you extract a Lotus Domino XML Language (.DXL) file, the message text and header information (*To*, *From*, *Sent*, and so on) is extracted to a text file.

NOTE: To prevent header information from being extracted, see [Exclude Metadata from the Extracted Text File, on page 57](#).

You can make sure that dates and times extracted from Lotus Domino .DXL files are displayed in a uniform format.

To extract custom date/time formats

- In the `formats_e.ini` file, set the `DateTimeFormat` option in the `[dxlsr]` section. For example:

```
[dxlsr]
DateTimeFormat=%m/%d/%Y %I:%M:%S %p
```

In this example, dates and times are extracted in the following format:

02/11/2003 11:36:09 AM

The format arguments are the same as those for the `strftime()` function. See <http://msdn.microsoft.com/en-us/library/fe06s4ak%28VS.71%29.aspx> for more information.

Extract .DXL Files to HTML

You can use the file extraction API to process .DXL files with an XSLT engine. The XSLT engine then transforms the extracted .DXL to .mail HTML files.

To extract .DXL files to HTML

- Set the following options in the `formats_e.ini` file:

```
[nsfsr]
ExportDXL=1
ExportDXL_PureXML=1
[dxlsr]
LNDParser=2
```

Extract Subfiles from Lotus Notes Database Files

A Lotus Notes database is a single file that contains multiple documents called *notes*. Notes include design notes (such as forms, views, folders, navigators, outlines, pages, framesets, agents, and resources), data document notes, profile document notes, access control list notes, and collection (index) notes. KeyView can extract text items, attachments, and OLE objects from *data document notes* only. Data document notes include emails, journal entries, discussion threads, documents (Microsoft Office and Lotus SmartSuite), and so on.

All components of a note are prefixed by field names such as "SendTo:", "Subject:", and "Body:". When a note is extracted, the field names are not included in the extracted output; only the field values are extracted.

When a mail message in an NSF file is extracted to disk, the body text and header information (such as the values from the `SendTo`, `From`, and `DeliveredDate` fields) in each message is extracted to a text file. (If you do not want the header information to appear in the message text file, see [Exclude Metadata from the Extracted Text File, on page 57.](#))

NOTE: The Lotus Notes Database (NSF) reader is an advanced feature and is sold and licensed separately. To enable this reader in a KeyView SDK, you must obtain the appropriate license key from Micro Focus.

System Requirements

The NSF format is proprietary. Therefore, KeyView accesses NSF files indirectly by using the Lotus Notes API. Because the NSF reader relies on functionality in Lotus Notes, a Lotus Notes client or Lotus Domino server must be installed and configured on the same machine on which the application converting NSF files is installed. On UNIX and Linux, the Lotus Domino server is required. On Windows, the Lotus Notes client or Lotus Domino server is required.

KeyView supports the following Lotus Notes clients and Domino servers:

- Lotus Notes 6.5.1
- Lotus Domino 6.5.1

KeyView supports NSF files on the same platforms supported by Lotus Notes and Lotus Domino:

- Windows XP x86 (Service Pack 1 and 2)
- Windows 2000 x86 (Service Pack 2)
- Solaris 8.0 and 9.0 (built on Solaris 8.0)
- Red Hat Enterprise Linux AS 3.0 (x86)
- SuSE Linux Enterprise Server 8 and 9 (x86)
- IBM AIX 5.1, 5L version 5.2

Installation and Configuration

Before KeyView can convert NSF files, you must set up the Lotus Notes client or Lotus Domino server. Full configuration is not required. The following steps outline the minimal setup for NSF conversion.

Windows

1. Install the Lotus Notes client or Lotus Domino server. You do not need to configure the client or server.
2. Make sure that the `notes.ini` file is in the proper location.
 - If Lotus Notes is installed, the file should appear in the `install\lotus\notes` directory, where `install` is the installation directory.
 - If only Lotus Domino is installed, the file should appear in the `install\lotus\domino` directory, where `install` is the installation directory.

If the file does not exist, create an ASCII file named `notes.ini`, and add the following text:

```
[Notes]
```

3. Add the KeyView `bin` directory and the `install\lotus\notes` or `install\lotus\domino` directory to the `PATH` environment variable (the KeyView `bin` directory must be first in the path). Micro Focus recommends that you add the KeyView `bin` directory because the Lotus Notes or Domino server installation might contain older KeyView OEM libraries.

Solaris

1. Install Lotus Domino server. You do not need to configure the server.
2. Make sure that the file `notes.ini` is in the `install/lotus/notes/latest/sunspa` directory, where `install` is the directory where Lotus Notes is installed. If the file does not exist, create an ASCII file named `notes.ini`, and add the following text:

```
[Notes]
```

3. Add the *install/lotus/notes/latest/sunspa* directory to the PATH environment variable:

```
setenv PATH install/lotus/notes/latest/sunspa:$PATH
```

4. Add the *install/lotus/notes/latest/sunspa* and the KeyView bin directory to the LD_LIBRARY_PATH environment variable:

```
setenv LD_LIBRARY_PATH keyview_bin:install/lotus/notes/latest/sunspa:$LD_LIBRARY_PATH
```

where *keyview_bin* is the location of the KeyView bin directory. Micro Focus recommends that you add the KeyView bin directory because the Lotus Notes installation might contain older KeyView OEM libraries.

AIX 5.x

1. Install the *bos.iocp.rte* file set if it is not already installed, and reboot the machine. See the Lotus Domino server documentation for more information.
2. Install Lotus Domino server. You do not need to configure the server.
3. Make sure that the file *notes.ini* is in the *install/lotus/notes/latest/ibmpow* directory, where *install* is the directory where Lotus Notes is installed. If the file does not exist, create an ASCII file named *notes.ini*, and add the following text:

```
[Notes]
```

4. Add the *install/lotus/notes/latest/ibmpow* directory to the PATH environment variable:

```
setenv PATH install/lotus/notes/latest/ibmpow:$PATH
```

5. Add the *install/lotus/notes/latest/ibmpow* and the KeyView bin directory to the LIBPATH environment variable:

```
setenv LIBPATH keyview_bin:install/lotus/notes/latest/ibmpow:$LIBPATH
```

where *keyview_bin* is the location of the KeyView bin directory. Micro Focus recommends that you add the KeyView bin directory because the Lotus Notes installation might contain older KeyView OEM libraries.

Linux

1. Install Lotus Domino server. You do not need to configure the server.
2. Make sure the file *notes.ini* is in the *install/lotus/notes/latest/linux* directory, where *install* is the directory where Lotus Notes is installed. If the file does not exist, create an ASCII file named *notes.ini*, and add the following text:

```
[Notes]
```

3. Add the *install/lotus/notes/latest/linux* directory to the PATH environment variable:

```
setenv PATH install/lotus/notes/latest/linux:$PATH
```

4. Add the *install/lotus/notes/latest/linux* and the KeyView bin directory to the LD_LIBRARY_PATH environment variable:

```
setenv LD_LIBRARY_PATH keyview_bin:install/lotus/notes/latest/linux:$LD_
LIBRARY_PATH
```

where *keyview_bin* is the location of the KeyView bin directory. Micro Focus recommends that you add the KeyView bin directory because the Lotus Notes installation might contain older KeyView OEM libraries.

Open Secured NSF Files

KeyView enables you to specify a user ID file and password to use to open a secured NSF file for extraction.

Format Note Subfiles

The KeyView NSF reader uses XML templates to format note subfiles. You can customize the templates to approximate the look and feel of the original notes as closely as possible. For more information, see [Extract and Format Lotus Notes Subfiles, on page 216](#).

Extract Subfiles from PDF Files

KeyView can extract document-level and page-level attachments from a PDF document. Document-level attachments are added by using the **Attach A File** tool, and can include links to or from the parent document or to other file attachments. Page-level attachments are added as comments by using various tools. Page-level or comment attachments display the File Attachment icon or the Speaker icon on the page where they are located.

When a PDF file is extracted to disk, the PDF file is extracted to a directory and the PDF's attachments are saved in their native format to the same directory as the original PDF file.

Improve Performance for PDFs with Many Small Images

To improve performance when processing PDF files that contain many small images, you can choose to ignore images unless they exceed a minimum width and/or height. If an image is smaller than the minimum width or height, KeyView does not extract the image.

For example, to ignore images that are less than 16 pixels wide or less than 16 pixels in height, add the following to the [pdf_flags] section of the *formats_e.ini* file:

```
[pdf_flags]
process_images_with_min_width=16
process_images_with_min_height=16
```

Extract Embedded OLE Objects

Embedded OLE objects can be converted in two ways:

- Using the File Extraction API, the OLE object is first extracted from the main file and saved to disk. It can then be converted by making a separate conversion call.
- Using the HTML Export API, the main file is converted to HTML and the OLE object is converted to a graphics file that is referenced in the HTML file.

The File Extraction API can extract embedded OLE objects from the following types of documents:

- Lotus Notes (DXL)
- Microsoft Excel
- Microsoft Word
- Microsoft PowerPoint
- Microsoft Outlook
- Microsoft Visio
- Microsoft Project
- OASIS Open Document
- Rich Text Format (RTF)

When an embedded OLE object is extracted from its parent file, the location of the embedded file in the original document is not available. The parent and child are extracted as separate files.

Extract Subfiles from ZIP Files

You can extract ZIP files that are not password-protected by using the general method (see [Extract Subfiles, on page 48](#)). However, some ZIP files use password protection, in which case you must use a different method to enter the required credentials.

Default File Names for Extracted Subfiles

When a file name is not specified in the call to `extExtractSubFile`, in some cases, a default file name is applied to the extracted subfile.

Default File Name for Mail Formats

To avoid naming conflicts and problems with long file names, KeyView applies its own names to the extracted mail folders and mail items when a name is not supplied in the call to `extExtractSubFile`. A non-mail attachment retains its original file name and extension.

When the contents of a mail store or the message body of a mail message are extracted, the extracted file names might include the following:

- The first valid eight characters of the original folder name or "Subject" line of the mail message. If the "Subject" line is empty, the characters `kvext` are used, where `ext` is the format's extension. For example, the characters would be `"kvmsg"` for MSG, and `"kvnsf"` for NSF.

The following special characters are considered invalid and are ignored:

any non-printing character with a value less than 0x1F

angle brackets (< >) double quotation mark (")

asterisk (*) forward slash (/)

back slash (\) pipe (|)

colon (:) question mark (?)

For notes, the file name is derived from the first 24 characters of the note text. For contact entries, the file name is derived from the full name of the contact.

- The characters `_kvn`, where `n` is an integer incremented from 0 for each extracted item.
- One of the following extensions:

| Type | File Extension |
|----------------------|----------------|
| email message | .mail |
| calendar appointment | .cal |
| contact entry | .cont |
| task entry | .task |
| note | .note |
| journal entry | .jrn1 |
| distribution list | .dist |
| posting note | .post |

If the type cannot be determined for an MSG or PST file, the file is given a `.mail` extension.

If the type cannot be determined for an NSF file, the file is given a `.tmp` extension.

For example, an MSG mail message with the subject line "RE: Product roadmap" that contains the Microsoft Excel attachment `release_schedule.xls` is extracted as

```
RE produ_kv0.mail
```

```
release_schedule.xls
```

If an extracted message contains an embedded OLE object or any attachment that does not have a name, the object or attachment is extracted as `_kv#.tmp`.

Default File Name for Embedded OLE Objects

KeyView can apply a default name to an extracted embedded OLE object when a name is not supplied in the call to `extExtractSubFile`. When an embedded OLE object is extracted, the extracted file name might include the following:

- The first valid eight characters of the main file. The following special characters are considered invalid and are ignored:

any non-printing character with a value less than 0x1F

angle brackets (< >) double quotation mark (")

asterisk (*) forward slash (/)

back slash (\) pipe (|)

colon (:) question mark (?)

- The characters `_kvn`, where `n` is an integer incremented from 0 for each extracted object.
- If KeyView can determine the embedded OLE is a Microsoft Office document, the original extension is used. If the file type cannot be determined, the file is given a `.tmp` extension.

For example, let us say a Microsoft Word document (`sales_quarterly.doc`) contains two embedded OLE objects: a Microsoft Excel file called `west_region.xls`, and a bitmap created in the Word document. The embedded objects would be extracted as

`sales_qu_kv0.xls`

`sales_qu_kv1.tmp`

Exclude Japanese Guide Text

This option prevents output of Japanese phonetic guide text when Microsoft Excel (`.xlsx`) files are processed.

To prevent output of Japanese phonetic guide text

- Set `NoPhoneticGuides` to `TRUE` in the `formats_e.ini` file:

```
[Options]
NoPhoneticGuides=TRUE
```

You can also enable this option programatically when filtering by passing `KVFLT_NOPHONETICGUIDES` to `fpFilterConfig`.

Chapter 4: Use the HTML Export API

This section describes how to perform some basic tasks using the HTML Export API.

| | |
|---|-----|
| • Extract Metadata | 69 |
| • Extract File Format Information | 74 |
| • Convert Character Sets | 75 |
| • Map Styles | 79 |
| • Use Style Sheets | 82 |
| • Display Vector Graphics on UNIX and Linux | 83 |
| • Search and Highlight Terms | 84 |
| • Convert Revision Tracking Information | 84 |
| • Extract Text from Text Boxes | 85 |
| • Convert PDF Files | 85 |
| • Convert Spreadsheet Files | 94 |
| • Convert XML Files | 97 |
| • Error Messages | 102 |
| • Show Hidden Data | 105 |
| • Exclude Japanese Guide Text | 107 |
| • Source Code Identification | 107 |

Extract Metadata

When a file format supports metadata, KeyView can extract and process that information. Metadata includes document information fields such as title, author, creation date, and file size. Depending on the file's format, metadata is referred to in a number of ways: for example, "summary information," "OLE summary information," "file information," and "document properties."

The metadata in mail formats (MSG and EML) and mail stores (PST, NSF, and MBX) is extracted differently than other formats. For information on extracting metadata from these formats, see [Extract Mail Metadata, on page 52](#).

NOTE: KeyView can extract metadata from a document only if metadata is defined in the document, and the document reader can extract metadata for the file format. The section [Supported Formats, on page 121](#) lists the file formats for which metadata can be extracted. KeyView does not generate metadata automatically from the document contents.

Extract Metadata Using the API

You can extract the metadata at the API level. The API extracts all valid metadata fields that exists in the file.

To extract metadata using the Java API

1. Set the input source using the `setInputSource` method.
2. Call the `getSummaryInfo()` method of the `Export` object to retrieve an object of the `SummaryInfo` class.
3. Use the methods of the `SummaryInfo` object to retrieve the metadata information.

The `HtmlTest` sample program demonstrates how to extract metadata through the Java API.
See [HtmlTest](#), on page 112.

Example

```
SummaryInfo[] sinfo = objHtmlExport.getSummaryInfo();

if(sinfo != null)
{
    System.out.println("\nSummary info has been extracted.");
    fos_sum = new FileOutputStream(summaryOutFile);
    DataOutputStream dos_sum = new DataOutputStream(fos_sum);

    for(int i=0; i<sinfo.length; i++)
    {
        if(sinfo[i].getElementName() != null)
        {
            dos_sum.writeBytes("Element name: " + sinfo[i].getElementName() + "\n");
            dos_sum.writeBytes("Element type: " + sinfo[i].getSumInfoType() + "\n");

            if(sinfo[i].getIsValid() == true)
            {
                if(sinfo[i].isDateTimeType())
                {
                    dos_sum.writeBytes("Date/time: ");
                    dos_sum.writeBytes(sinfo[i].getDateTime())
                }
                else
                {
                    byte[] data = sinfo[i].getData();

                    if(data != null)
                    {
                        dos_sum.writeBytes("Element data: ");
                        dos_sum.write(data);
                    }
                }
            }
            dos_sum.writeBytes("\n\n");
        }
    }
}
```

```
dos_sum.close();  
fos_sum.close();  
}  
sinfo = null;
```

The `SummaryInfo` class stores the metadata extraction results. After calling the `HtmlExport.getSummaryInfo()` method, call the get methods provided by each instance of this class to extract metadata.

The following describes each get method:

| | |
|-------------------------------|---|
| <code>getElementName()</code> | This method gets the name of the metadata element. |
| <code>getSumInfoType()</code> | <p>This method specifies the data type of the metadata element. The possible types are:</p> <ul style="list-style-type: none">• <code>KV_String</code>• <code>KV_Int4</code>• <code>KV_DateTime</code>• <code>KV_ClipBoard</code>• <code>KV_Bool</code>• <code>KV_Unicode</code>• <code>KV_IEEE8</code>• <code>KV_Other</code> <p>If type is <code>KV_Bool</code>, data contains either <code>TRUE</code> or <code>FALSE</code>.</p> <p><code>KV_DateTime</code> and <code>KV_IEEE8</code> point to an 8-byte value.</p> |
| <code>getIsValid()</code> | This method specifies whether the data value is present in the document. <code>TRUE</code> specifies that the value is valid. For example, if the "Title" element was not populated in the document, <code>getIsValid</code> would return <code>FALSE</code> . |
| <code>isDateTimeType()</code> | This method determines whether the metadata element is of date/time type. |
| <code>getDateTime()</code> | This method gets the date and time in the form of a string. If the metadata element is of date/time type, call this method to get the date and time in the form of a string, for example "Wed Jun 30 21:49:08 1993" or "135 Minutes". |
| <code>getData()</code> | <p>This method gets the content of the element.</p> <p>If the metadata field is a date and time, the type is a 64-bit value representing the number of 100-nanosecond intervals since January 1, 1601.</p> <p>You can also use the <code>isDateTimeType()</code> method to determine whether a metadata element is of date/time type, and then use the <code>getDateTime()</code> method to obtain the date/time in the form of a string.</p> |

Extract Metadata Using a Template File

When using a template file, KeyView recognizes two types of metadata: *standard* and *non-standard*. Standard metadata includes fields, such as Title, Author, and Subject. The standard fields are enumerated from 1 to 41 in `KVSumType` in the `install\htmlexport\include\kvtypes.h` header file. Non-standard metadata includes any field not listed from 1 to 41 in `KVSumType`, such as user-defined fields (for example, custom property fields in Microsoft Word documents), or fields that are unique to a particular file type (for example, "Artist" or "Genre" fields in MP3 files). Enumerated types 42 and greater are reserved for non-standard metadata.

To extract metadata using a template file

1. Insert metadata tokens in a member of the `KVHTMLTemplateEx` section of the template file. This defines the point at which the metadata appears in the HTML output.
2. If you are using the `$USERSUMMARY` or `$SUMMARY` token, define the `szUserSummary` member of the `KVHTMLTemplateEx` section of the template file. This determines the markup and tokens generated when these metadata tokens are processed.

You can use the following metadata tokens in the template files:

| Token | Description |
|----------------------------|--|
| <code>\$SUMMARYNN</code> | This token inserts the data from a specified metadata field. <code>NN</code> is a number from 00 through 42 enumerated in <code>KVSumType</code> in <code>kvtypes.h</code> . |
| <code>\$SUMMARY</code> | This token inserts the data from valid metadata fields in the range of 0 to 42 using the markup provided in <code>pszUserSummary</code> . |
| <code>\$USERSUMMARY</code> | This token inserts the data from every valid non-standard metadata field using the markup provided in <code>pszUserSummary</code> . |
| <code>\$CONTENT</code> | This token inserts the content of the metadata field specified by the <code>\$NAME</code> token. |
| <code>\$NAME</code> | This token inserts the name of a the metadata field, such as "Title," "Author," or "Subject." |

Depending on the mark-up in `szUserSummary`, the extracted metadata might not appear in the browser when the HTML file is displayed, but might appear in the output file. Most of the KeyView-supplied template files extract standard metadata from a document, and include it in the output HTML. However, they do not display the metadata in a browser.

Examples

`$SUMMARYNN`

The following markup displays the contents of the "Title" field at the top of the main HTML file:

```
szMainTop=<em><strong>$SUMMARY01</strong></em>
```

In `KVSumType`, 01 is the enumerated value for the "Title" metadata field.

\$SUMMARY

The following markup extracts all standard fields, and includes them in the first heading level 1 HTML block:

```
szFirstH1Start=$SUMMARY
szUserSummary=<meta name="$NAME" content="$CONTENT" />
```

This example extracts the field name (\$NAME) and field content (\$CONTENT) for standard metadata from a document, and includes it at the beginning of the first heading level 1 HTML block. However, it does not display the metadata in the browser. The HTML output might look like this:

```
<meta name="CodePage" content="1252" />
<meta name="Title" content="My design document" />
<meta name="Subject" content="design specifications" />
<meta name="Author" content="John Doe" />
<meta name="Keywords" content="" />
<meta name="Comments" content="" />
<meta name="Template" content="Normal.dot" />
<meta name="LastAuthor" content="lchapman" />
<meta name="RevNumber" content="6" />
<meta name="EditTime" content="01/01/1601, 0:08" />
<meta name="LastPrinted" content="14/01/2002, 14:06" />
<meta name="Create_DTM" content="27/08/2003, 10:31" />
<meta name="LastSave_DTM" content="29/08/2003, 14:07" />
<meta name="PageCount" content="1" />
<meta name="WordCount" content="4062" />
<meta name="CharCount" content="23159" />
<meta name="AppName" content="Microsoft Word 9.0" />
<meta name="Security" content="0" />
<meta name="Category" content="software" />
<meta name="LineCount" content="192" />
<meta name="ParCount" content="46" />
<meta name="Manager" content="" />
<meta name="Company" content="Autonomy" />
```

To display the metadata in a browser, use the following markup in szUserSummary:

```
<hr>name="$NAME" content="$CONTENT"<br>
```

\$USERSUMMARY

The following markup extracts non-standard fields, and includes them at the bottom of the main HTML file:

```
szMainBottom=$USERSUMMARY
szUserSummary=<meta name="$NAME" content="$CONTENT" />
```

This example extracts the field name (\$NAME) and field content (\$CONTENT) for non-standard metadata from a document, and includes it at the bottom of the main HTML file. However, it does not display the metadata in the browser. The HTML output might look like this:

```
meta name="Telephone number" content="444-111-2222"  
meta name="Recorded date" content="07/03/2003, 23:00"  
meta name="Source" content="TRUE"  
meta name="my property" content="reserved"
```

To display the metadata in a browser, use the following markup in `szUserSummary`:

```
<hr>name="$NAME" content="$CONTENT"<br>
```

Extract File Format Information

Export can detect a file's format, and report the information to the API, which in turn reports the information to the developer's application. This feature enables you to apply customized conversion settings based on a file's format. See [File Format Detection, on page 232](#) for more information on format detection.

To extract file format information

1. Set the input source using the `setInputSource` method.
2. Call the `getAutoDetectInfo` method of the `Export` object to retrieve an object of the `AutoDetectInfo` class.
3. Use the methods of the `AutoDetectInfo` object to retrieve the format information.

The `HtmlTest` sample program demonstrates how to extract format information through the Java API. See [HtmlTest, on page 112](#).

Example

```
AutoDetectInfo adinfo = objHtmlExport.getAutoDetectInfo();  
if(adinfo != null)  
{  
    outf_format = new File(docFormatOutFile);  
    fos_format = new FileOutputStream(outf_format);  
    DataOutputStream dos_format = new DataOutputStream(fos_format);  
    dos_format.writeBytes("Auto-detection result: \n");  
    dos_format.writeBytes("\nCharacter set: " + adinfo.getCharacterSet());  
    dos_format.writeBytes("\nDocument class: " + adinfo.getDocumentClass());  
    dos_format.writeBytes("\nDocument format: " + adinfo.getDocumentFormat());  
    dos_format.writeBytes("\nFormat version: " + adinfo.getVersion());  
    dos_format.writeBytes("\nOther attributes:");  
    if(adinfo.isAppleDoubleEncoded())  
    {  
        dos_format.writeBytes("\nApple double encoded.");  
    }  
    if(adinfo.isAppleSingleEncoded())  
    {  
        dos_format.writeBytes("\nApple single encoded.");  
    }  
}
```

```
if(adinfo.isEncrypted())
{
    dos_format.writeBytes("\nEncrypted.");
}
if(adinfo.isMacBinaryEncoded())
{
    dos_format.writeBytes("\nMac binary encoded.");
}
if(adinfo.isWangGDLEncoded())
{
    dos_format.writeBytes("\nWang GDL encoded.");
}
dos_format.close();
fos_format.close();
adinfo = null;
```

Convert Character Sets

Export enables you to control the character set of both the input and the output text. This is accomplished by either

- setting the source, the target character set, or both, in the API
- basing the input/output on the character set of the document (if the document character set is stored in the document and can be determined by the document reader)

The character sets are defined as constants in the Export class. Not all character sets can be used to specify the target character set. See [Coded Character Sets, on page 210](#) for a list of character sets that can be used as a target character set.

Determine the Character Set of the Output Text

To determine the output character set of a converted document, Export considers the following:

- Whether the reader can extract the character set from the document. This depends on whether the file format can provide character set information and whether the document actually contains character set information.

The section [Supported Formats, on page 121](#) indicates the file formats for which character set information can be extracted. If character set information cannot be determined for your document type, you must set the source, the target character set, or both in the API.

- Whether the source character set is set in the API.

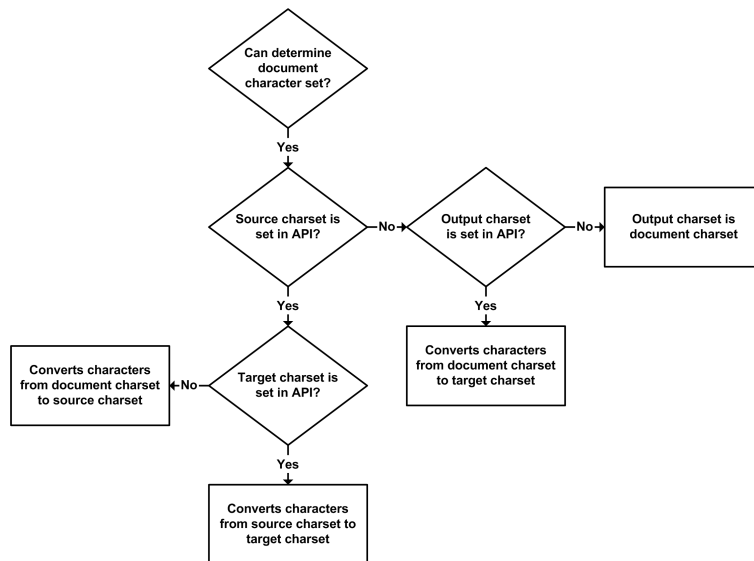
NOTE: To set the source character set, you must specify a character set *and* set the parameter `setForceSourceCharSet` to `TRUE`.

- Whether the target character set is set in the API.

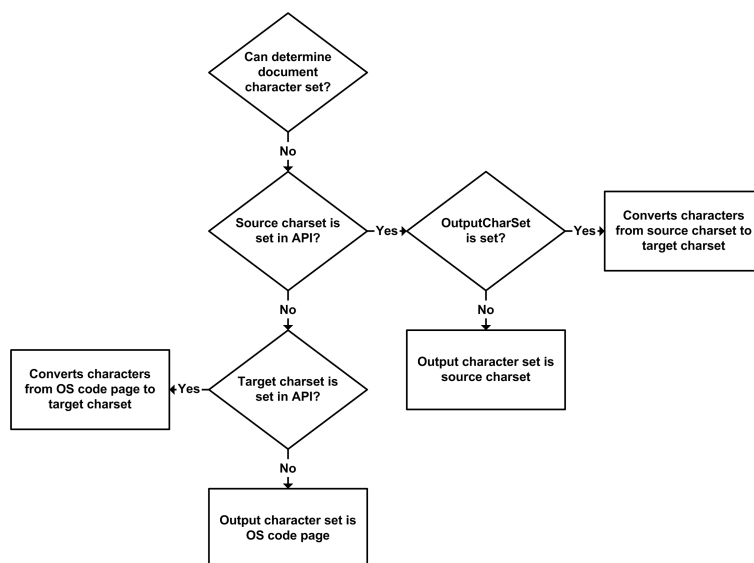
NOTE: To set the target character set, you must specify a character set *and* set the parameter `setForceOutputCharSet` to `TRUE`.

Guidelines for Character Set Conversion

The following diagram shows how the output character set is determined when the document character set can be determined.



The following diagram shows how the output character set is determined when the document character set cannot be determined.



Examples of Character Set Conversion

The examples below demonstrate possible configurations for mapping character sets and the expected output for each scenario.

Document Character Set Can be Determined

For the example in the following table, the document is an RTF file. The section [Word Processing Formats, on page 145](#) indicates that the document character set *can* be obtained from this file type. The document character set is Traditional Chinese (BIG5).

Document Character Set Can be Determined

| Source charset set | Target charset set | Output charset |
|--------------------|--------------------|--|
| KVCS_GB | KVCS_UTF8 | KVCS_UTF8 Converts GB (Simplified Chinese) to UTF-8. The output character set is the target character set specified in the API. |
| KVCS_GB | -- | KVCS_GB Converts BIG5 to GB (Simplified Chinese). The output character set is the source character set specified in the API. |
| -- | KVCS_UTF8 | KVCS_UTF8 Converts BIG5 to UTF-8. The output character set is the target character set specified in the API. |
| -- | -- | KVCS_BIG5 The output character set is the document character set. No conversion. |

Document Character Set Cannot be Determined

For the example in the following table, the document is an ASCII file. The section [Word Processing Formats, on page 145](#) indicates that the document character set *cannot* be obtained from this file type. The document's source character set is KVCS_1251.

Document Character Set Cannot be Determined

| Source charset set | Target charset set | Output charset |
|--------------------|--------------------|----------------|
| KVCS_1251 | KVCS_UTF8 | KVCS_UTF8 |

Document Character Set Cannot be Determined, continued

| Source charset set | Target charset set | Output charset |
|--------------------|--------------------|--|
| | | Converts KVCS_1252 to KVCS_UTF8. The output character set is the target character set specified in the API. |
| KVCS_1252 | KVCS_UNKNOWN | KVCS_1252 The output character set is the source character set specified in the API because KVCS_UNKNOWN cannot be used. No conversion. |
| KVCS_1252 | -- | KVCS_1252 The output character set is the source character set specified in the API. No conversion. |
| -- | KVCS_1252 | KVCS_1252 Converts OS code page to KVCS_1252. The output character set is the target character set specified in the API. |
| -- | -- | The output character set is OS code page. No conversion. |

Set the Character Set During Conversion

You can convert the character set of a file at the time the file is converted.

To specify the source character set, use the `setSourceCharSet` method of the `OptionInfo` object and set `setForceSourceCharSet` to `TRUE`.

To specify the target character set, use the `setOutputCharSet` method of the `OptionInfo` object and set `setForceOutputCharSet` to `TRUE`.

Set the Character Set During File Extraction from a Container

You can convert the character set of a container subfile at the time the subfile is extracted from the container and before it is converted to HTML. This is most often used to set the character set of a mail message's body text. See [Use the File Extraction API, on page 47](#).

To specify the source and target character set of a subfile

1. Use the methods of the `ExtSubFileExtractConfig` object to set the source and target character set.
2. Call the `extExtractSubFile` method of the `Export` object and pass in the `ExtSubFileExtractConfig` object.

Map Styles

Export can map paragraph and character styles in any word processing format that contains styles (such as Microsoft Word, RTF, or Folio Flat File) to user-defined markup. This feature is useful for shaping the look of the HTML output, or for generating user-defined metadata (including using XML tags) for indexing, searching, and navigation. With this feature, you can redact (hide) text in the source document, delete content, or change the overall structure of the output. You can also embed style sheet styles in the output defined in the HTML.

To enable style mapping, you must indicate which paragraph and/or character styles are to be mapped, and define the starting and ending markup to be included in the HTML output. For example, if the source Microsoft Word document contains the character style "Recipe," and the content of the style in Microsoft Word is "Brownies," you can specify that the starting markup be `<recipe>` and the ending markup `</recipe>`. This would result in the output HTML containing: `<recipe>Brownies</recipe>`.

You can also use style mapping to control the look of the HTML output either by using a Cascading Style Sheet (CSS) or by defining the style directly in the starting markup. For example, if a Word document contains the paragraph style "Colorful", you can have markup of the form `<p><div class="rainbow">` inserted at the front of the paragraph and markup of the form `</div></p>` inserted at the end of the paragraph. "Rainbow" is a CSS style defined in an externally provided CSS file referenced at the top of the HTML output.

Style mapping is enabled in the `wordstyle.ini` template file. The HTML Export Getting Started page demonstrates the output resulting from a conversion using `wordstyle.ini`. The Getting Started page, named `htmstart.html`, is in the `install\htmlexport\docs` directory, where `install` is the path name of the Export installation directory. The source documents used in the page are in the directory `install\testdocs`.

NOTE: When the user-defined markup in `KVStyle` conflicts with other markup generated by HTML Export, the user-defined markup takes precedence.

Use the Java API

To map styles using the Java API

1. Create an instance of the `StyleMapping` class. Using the object's methods, specify the style mapping information. The information includes
 - the markup to be added to the beginning and end of a paragraph or character style.
 - the name of the word processing style (for example, "Heading 1") to which style mapping applies. Style names are case sensitive.
 - the flag which defines instructions on how to process the content associated with a paragraph or character style. The flags are defined as constants in the `Export` class and listed in [Flags for Defining Styles, on page 81](#).
2. Call the `setStyleMapping()` method of the `Export` object and pass the `StyleMapping` object.

Use a Template file

To map styles using a template file

1. Use the `KVStyle` parameter to specify how many styles are being mapped. For example, if there are nine mapped heading levels, add the following:

```
[KVStyle]
NumStyles=9
```

2. For each style, there must be a `[StyleX]` entry that contains the markup that appears at the start and end of the defined style. For example, in the `wordstyle.ini` sample file, the first heading level is defined as follows:

```
[Style1]
StyleName=Colorful
MarkupStart=<div class="colorful">
MarkupEnd=<!-- end of colorful --></div>
```

3. For each style, define the flag that applies. Flags define instructions on how to process the content associated with a paragraph or character style. They are defined as constants in the Export class and listed in [Flags for Defining Styles, on the next page](#).

```
Flags=KVSTYLE_HEADING3
```

4. A finished entry in a template file could look like this:

```
[KVStyle]
NumStyles=3

[Style1]
StyleName=Colorful
MarkupStart=<div class="Colorful">
MarkupEnd=<!-- End of Colorful --></div>
Flags=KVSTYLE_HEADING1

[Style2]
StyleName=RedactPara
MarkupStart=<div class="RedactPara">
MarkupEnd=<!-- End of RedactPara --></div>
Flags=KVSTYLE_REDACT

[Style3]
StyleName=Code
MarkupStart=<pre>
MarkupEnd=<!-- End of Code --></pre>
Flags=KVSTYLE_PRE
```

Flags for Defining Styles

| Flag | Description |
|---------------------------------|--|
| KVSTYLE_PRE | The KVSTYLE_PRE flag specifies that white space should be preserved (treated as characters, not word separators), and that mode changes, such as changes in font size within a paragraph, should be ignored. This allows the tags <code><pre></code> and <code></pre></code> to be used. |
| KVSTYLE_HEADING[1-6] | <p>The KVSTYLE_HEADING[1-6] flags specify that a specific style is to be detected and processed as a heading. Heading flags are exclusive. This means that a style cannot be processed as both h1 and h2.</p> <p>By default, Export maps the heading style "Heading 1" to <code><h1></h1></code>, and so on, for heading levels 1 through 6. If you use style mappings, the default mapping is overridden. Therefore, you must supply markup for all heading levels. HTML Export uses heading levels to define the overall structure of the HTML output.</p> |
| KVSTYLE_ORDERLIST | The KVSTYLE_ORDERLIST flag specifies that the style should be tagged as an ordered list. Currently not implemented. |
| KVSTYLE_UNORDEREDLIST | The KVSTYLE_UNORDEREDLIST flag specifies that the style should be tagged as an unordered list. Currently not implemented. |
| KVSTYLE_DELETECONTENT | The KVSTYLE_DELETECONTENT flag specifies that the content associated with the style tag should be deleted from the output. |
| KVSTYLE_ONCONSECUTIVEPARAGRAPHS | The KVSTYLE_ONCONSECUTIVEPARAGRAPHS flag specifies that the style should be applied to consecutive paragraphs of the document. If this flag is used, and two or more paragraphs require the same style, the opening and closing tags that normally appear between each paragraph are not generated. |
| KVSTYLE_REDACT | The KVSTYLE_REDACT flag is used to hide sensitive or confidential information in the source document. It specifies that the text associated with the style tag should be replaced in the HTML output with a selected character. The default replacement character is "X", but you can specify a different replacement character by using the <code>setRedact(java.lang.Byte b)</code> method of the <code>OptionInfo</code> class. |
| KVSTYLE_DEFAULT | The KVSTYLE_DEFAULT flag specifies that special processing should not be applied to the content. |

Use Style Sheets

You can use style sheets to define the overall layout and type specifications of the HTML output. HTML Export can write style sheet information to an external Cascading Style Sheet (CSS) file or read the information from an existing CSS file during the conversion. The formatting data can be stored either within the output HTML file (inline) or externally in a CSS file. Using an external style sheet makes the HTML output significantly smaller and enables you to use the same style sheet for many conversions.

NOTE: Cascading style sheets can be used only with word processing documents.

To enable CSS formatting and output the generated formatting data within the output HTML stream, set the style sheet type to inline by using one of the following methods:

- Use the `setStyleSheetType` method of the `HtmlOptionInfo` class to set the option to `HtmlExport.STYLESHEET_CSS_INLINE`.
- Set the `eStyleSheetType` parameter in a template file to `CSS_INLINE`.

To enable CSS formatting and output the generated formatting data in an external CSS file that is referenced in the HTML output as a tag

1. Set the style sheet type to an external file by using one of the following methods:
 - Use the `setStyleSheetType` method of the `HtmlOptionInfo` class to set the type to `HtmlExport.STYLESHEET_CSS_TOFILE`.
 - Set the `eStyleSheetType` parameter in a template file to `CSS_TOFILE`.
2. In the template file, use the `$STYLESHEET` token to specify the URL of the style sheet in the HTML output. The external CSS file is referenced in the output HTML by a `LINK` statement of the form:

```
<LINK rel="STYLESHEET" href="CSS_file" type="text/css">
```
3. Call the `setExternalStyleFile` method of the `Export` object to set the path and file name of the style sheet file.
 - If the file does not exist in the specified location, it is created.
 - If the file exists, but is empty, CSS styles are written to the file.
 - If the file exists and is not empty, the conversion attempts to use the predefined styles, and appends any new styles that are required for the conversion.
 - If the file is not a CSS file or if it points to a non-existent directory, the `LINK` statement is written, but the style sheet information is added inline (`CSS_INLINE`), and an external CSS file is not generated.
 - If a style file has been set, you can disable it for the next convert call by calling the `setExternalStyleFile` method again and passing in an empty string.

Display Vector Graphics on UNIX and Linux

Export offers the option of rasterizing vector graphic content from source documents into a variety of graphics formats, including JPEG, PNG, WMF, and CGM. This solution is implemented with Windows Graphical Device Interface (GDI) code, and therefore is not portable to other platforms.

The output format of vector graphics is defined by using the `setOutputVectorGraphicType()` method in the `HtmlOptionInfo` class, and the options are defined as constants in the `Export` class.

To display vector graphics in presentation, word processing, and spreadsheet files on UNIX and Linux, Export can convert the files directly to JPEG by using a Java program named `kvraster.class`. This program uses the Java Abstract Windowing Toolkit (AWT). The AWT requires access to an X Server.

NOTE: If you are using KeyView 10.5.0.0 or Java 1.6, you do not have to set up an X Server; however, if you are using a version of KeyView lower than 10.4 with a version of Java lower than 1.6, you must set up an X Server.

To set up an X Server, do one of the following:

- Run a virtual X Server, such as the Xvfb utility. This utility is included in the X11R6 distribution or can be downloaded from the following site:

<http://www.x.org/Downloads.html>

For example, to run the Xvfb utility on a 512 Mb, Solaris 2.8 platform, follow these steps:

1. Start Xvfb at root:

```
/usr/X11R6/bin/Xvfb :1 -screen 0 1152x900x8 &
```

2. Set the display environment variable:

```
setenv DISPLAY:1.0
```

- Make an X display available to the Java runtime using the `DISPLAY` environment variable. No windows appear on the display. For example, set the `DISPLAY` environment variable as follows:

```
setenv DISPLAY computername:0.0
```

or

```
setenv DISPLAY ipaddress:0.0
```

After the X Server is set up, convert the file by following these steps:

1. Add the location of the JRE to the `PATH` environment variable.
2. Use one of the following methods to set the graphic type to JPEG:
 - Use the `setOutputVectorGraphicType` method of the `OptionInfo` class to set the type to `Export.GRAPHIC_TYPE_JPEG`.
 - Set the `OutputVectorGraphicType` parameter of the `defunix.ini` template file to `KVGFX_JPEG`.

3. Convert the document to HTML. The graphics in the document are converted to JPEG and stored in the output directory.

Search and Highlight Terms

Using the highlighting API, KeyView can find and highlight specified text strings in the HTML output. Only text strings that exactly match the search term are highlighted. For example, if the term *house* is specified, the string *house* would be highlighted in **house**, **houses**, and **housed**, but would not be highlighted in the term *housing*. You can define the text attributes used to highlight the text (such as bold, red, or underlined), and the text's target character set.

If a specified term contains HTML code, it is not found. For example, if the phrase *weekly schedule* was specified, the following string in the output HTML would not be found:

```
weekly <b>schedule</b>
```

If you specify multiple terms, and some terms are subsets of other terms, Micro Focus recommends that you specify the superset first. For example,

```
["North American car manufacturers" "car manufacturers" "car"]
```

To specify search terms using the Java API

1. Create an instance of the `HtmlHighlight` class and, using the object's methods, specify the list of terms, the highlighting attributes, and the case-sensitivity Boolean.
2. Call the `setHighlight()` method and pass the `HtmlHighlight` object.

Convert Revision Tracking Information

The revision tracking feature in applications—such as Microsoft Word's **Track Changes**—marks changes to a document (typically, strikethrough for deleted text and underline for inserted text) and tracks each change by reviewer name and date.

If revision tracking was enabled when changes were made to a document, you can configure Export to convert the deleted text and graphics and include revision tracking information in the HTML output. (The deleted content and revision tracking information is excluded from the HTML output by default.)

Content that was added to the document is identified by `<ins>` tags and is underlined when displayed in a browser. Content that was deleted from the document is identified by `` tags and is displayed with strikethrough formatting.

The `<ins>` and `` tags include `cite` and `datetime` attributes, which define the name of the reviewer who made the change and the date on which the change was made respectively. (The date is in ISO-8601 format: YYYY-MM-DDThh:mm:ss.) The tags also include a `title` attribute which displays the author and date information in a browser.

For example, the following markup is generated for inserted text:

```
<ins title="Inserted: JohnD, 2006-04-24T14:47:00" cite="mailto:JohnD"
datetime="2006-04-24T14:47:00">This text was added</ins> in a previous version.
```

This text is displayed in the browser as:

This text was added in a previous version.

When you hover the cursor over the underlined text in the browser, the text **"nserted: JohnD, 2006-04-24T14:47:00"** is displayed as a ToolTip.

The following markup is generated for deleted text:

```
<del title="Deleted: JohnD, 2006-04-24T14:56:00" cite="mailto:JohnD"
datetime="2006-04-24T14:56:00">This text was deleted</del> in a previous version.
```

This text is displayed in the browser as:

~~This text was deleted~~ in a previous version.

When you hover the cursor over the strikethrough text in the browser, the text **"Deleted: JohnD, 2006-04-24T14:56:00"** is displayed as a ToolTip.

NOTE: Whether the text is displayed with strikethrough or underline depends on the configuration and capabilities of the browser.

To convert deleted text and graphics and include revision tracking information, call the `includeRevisionMark` method. For example:

```
if(inclRevisionMark == true)
{
    objHtmlExport.includeRevisionMark();
}
```

To reset the flag and exclude deleted content and revision tracking information from the HTML output, call the `excludeRevisionMark` method. For example:

```
if(inclRevisionMark == false)
{
    objHtmlExport.excludeRevisionMark();
}
```

Extract Text from Text Boxes

By default, the contents of Microsoft Word text boxes are converted to graphics and exported. Alternatively, you can convert the contents of text boxes to text.

NOTE:
The extracted text box text is not formatted.

To enable text box extraction, add the following section to the `formats_e.ini` file:

```
[WordTextBoxOptions]
OutputText=true
```

Convert PDF Files

Export has special configuration options that allow greater control over the conversion of PDF files. These options can improve the fidelity and accuracy of the HTML output.

Use a Graphic-Based Reader

Two graphic-based PDF readers are available. The readers display PDFs by converting each page of the PDF to an image. If you do not want to redistribute the Acrobat Reader with your application, you can use a graphic-based reader instead.

The two readers support different features. Choose the appropriate reader depending on your requirements:

- The kppdfdr reader supports highlighting, annotation, and several other features but also has several graphical limitations.
- The kppdf2rdr reader produces high-fidelity raster images but is a viewer only and does not support highlighting or other features.

Use the kppdfdr Reader

The kppdfdr graphic-based reader has the following features:

- supports vector images
- supports rotation and scaling
- supports multibyte and bidirectional text
- allows you to search text in the output

The kppdfdr reader has the following limitations:

- Embedded fonts in a PDF file are not translated correctly. They are usually displayed using the question mark (?) replacement character.
- If an unsupported font is encountered during conversion, the default font, Times New Roman, is substituted.
- Supports 180 degree rotation only for raster images.
- Supports the following color spaces: DeviceRGB, DeviceGray, DeviceCMYK, CalGray, and CalRGB color spaces. Indexed color spaces are supported as long as they are used with a supported basic color space.
- Does not support hyperlinks.
- Does not extract summary information (metadata).

Use the kppdf2rdr Reader

The kppdf2rdr graphic-based reader produces high-fidelity raster images. However, it has the following limitations:

- Does not support anything beyond viewing, such as text searching.
- Does not support PDFs that contain XFA forms content.

Specify the Graphic-based Reader

By default, the Acrobat control is used to convert PDF documents. Use the following procedure to specify that one of the graphic-based readers be used to convert PDF documents.

To specify the graphic-based reader

1. Open the `formats_e.ini` file with a text editor. The file is installed in the root of the Windows directory.
2. In the `[HiFi]` section, set the following parameter to the graphic-based reader you want to use. Set one of the following values:
 - For the `kppdfrdr` reader:
`200=kppdfrdr.dll`
This is the default setting.
 - For the `kppdf2rdr` reader:
`200=kppdf2rdr.dll`
Set `CFG_SETHIFIPDF` field in the `HtmlExport` class.
3. Set `CFG_SETHIFIPDF` field in the `HtmlExport` class.

Convert PDF Files to Raster Images

Export allows you to convert each page of a PDF document to a raster image, providing a high-fidelity conversion of the document.

The output format depends on the value of `setOutputRasterGraphicType` in `HtmlOptionInfo`.

On UNIX and Linux, the conversion of PDFs to JPEG uses the Java program `kvraster.class`. This Java program requires some setup. See [Display Vector Graphics on UNIX and Linux, on page 83](#).

To specify the graphic-based reader for converting PDF documents

1. Specify the graphic-based reader you want to use.
2. Create an instance of the `ConfigOption` class. Set the `OptionType` argument to `CFG_SETHIFIPDF`, and the `OptionValue` argument to `1`.
3. Call the `setConfigOption` method and pass the `ConfigOption` object.
4. Call a convert method. See the Javadoc in the directory `install\javaapi\javadoc`, where `install` is the path name of the Export installation directory.

The `HtmlConvFileToFile` sample program demonstrates how to use the `setConfigOption()` method. See [HtmlConvFileToFile, on page 114](#).

Convert PDF Files to a Logical Reading Order

The PDF format is primarily designed for presentation and printing of brochures, magazines, forms, reports, and other materials with complex visual designs. Most PDF files do not contain the *logical structure* of the original document—the correct reading order, for example, and the presence and meaning of significant elements such as headers, footers, columns, tables, and so on.

KeyView can convert a PDF file either by using the file's internal unstructured paragraph flow, or by applying a structure to the paragraphs to reproduce the logical reading order of the visual page. Logical reading order enables KeyView to output PDF files that contain languages that read from right-to-left (such as Hebrew and Arabic) in the correct reading direction.

NOTE: The algorithm used to reproduce the reading order of a PDF page is based on common page layouts. The paragraph flow generated for PDFs with unique or complex page designs might not emulate the original reading order exactly.

For example, page design elements such as drop capitals, callouts that cross column boundaries, and significant changes in font size might disrupt the logical flow of the output text.

Logical Reading Order and Paragraph Direction

By default, KeyView produces an *unstructured* text stream for PDF files. This means that PDF paragraphs are extracted in the order in which they are stored in the file, not the order in which they appear on the visual page. For example, a three-column article could be output with the headers and the title at the end of the output file, and the second column extracted before the first column. Although this output does not represent a logical reading order, it accurately reflects the internal structure of the PDF.

You can configure KeyView to produce a *structured* text stream that flows in a specified direction. This means that PDF paragraphs are extracted in the order (logical reading order) and direction (left-to-right or right-to-left) in which they appear on the page.

The following paragraph direction options are available.

| Paragraph Direction Option | Description |
|----------------------------|---|
| Left-to-right | Paragraphs flow logically and read from left to right. You should specify this option when most of your documents are in a language that uses a left-to-right reading order, such as English or German. |
| Right-to-left | Paragraphs flow logically and read from right to left. You should specify this option when most of your documents are in a language that uses a right-to-left reading order, such as Hebrew or Arabic. |
| Dynamic | Paragraphs flow logically. The PDF reader determines the paragraph direction for each PDF page, and then sets the direction accordingly. This option is used when a paragraph direction is not specified. |

NOTE: Conversions might be slower when logical reading order is enabled. For optimal speed, use an unstructured paragraph flow.

The paragraph direction options control the direction of paragraphs on a page; they do not control the text direction in a paragraph. For example, a PDF file might contain English paragraphs in three columns that read from left to right, but 80% of the second paragraph contains Hebrew characters. If you enable the left-to-right logical reading order, the paragraphs are ordered logically in the output—title paragraph, then paragraph 1, 2, 3, and so on—and flow from the top left of the first column to the bottom right of the third column. However, the *text* direction of the second paragraph is determined independently of the page by the PDF reader, and is output from right to left.

NOTE: Extraction of metadata is not affected by the paragraph direction setting. The characters and words in metadata fields are extracted in the correct reading direction regardless of whether logical reading order is enabled.

Enable Logical Reading Order

You can enable logical reading order by using either the API or the `formats_e.ini` file. Setting the direction in the API overrides the setting in the `formats_e.ini` file.

Use the Java API

To enable PDF logical reading order in the Java API

1. Use the `setPDFLogicalOrder(int orderFlag)` method of the `HtmlExport` object, and set the `orderFlag` argument to one of the following flags.

| Flag | Description |
|------------------------|---|
| PDF_LOGICAL_ORDER_LTR | Logical reading order and left-to-right paragraph direction |
| PDF_LOGICAL_ORDER_RTL | Logical reading order and right-to-left paragraph direction |
| PDF_LOGICAL_ORDER_AUTO | Logical reading order. The PDF reader determines the paragraph direction for each PDF page, and then sets the direction accordingly. This option is used when a paragraph direction is not specified. |
| PDF_LOGICAL_ORDER_RAW | Unstructured paragraph flow. This is the default behavior. Set this flag if logical reading order is enabled, and you want to return to an unstructured paragraph flow. |

For example,

```
objHTMLExport.setPDFLogicalOrder(Export.PDF_LOGICAL_ORDER_RTL);
```

Use the formats_e.ini File

The `formats_e.ini` file is in the `install\OS\bin` directory, where `install` is the path name of the Export installation directory and `OS` is the name of the operating system.

To enable logical reading order by using the formats_e.ini file

1. Change the PDF reader entry in the `[Formats]` section of the `formats_e.ini` file as follows:

```
[Formats]
200=1pdf
```

2. Optionally, add the following section to the end of the `formats_e.ini` file:

```
[pdf_flags]
pdf_direction=paragraph_direction
```

where `paragraph_direction` is one of the following.

| Flag | Description |
|---------------|--|
| LPDF_ LTR | Left-to-right paragraph direction |
| LPDF_ RTL | Right-to-left paragraph direction |
| LPDF_ AUTO | The PDF reader determines the paragraph direction for each PDF page, and then sets the direction accordingly. This option is used when a paragraph direction is not specified. |
| LPDF_ RAW | Unstructured paragraph flow. This is the default behavior. Set this flag if logical reading order is enabled, and you want to return to an unstructured paragraph flow. |

Generate a Table of Contents from PDF Bookmarks

When you use the basic reader (`pdfsr`) to convert PDF files to HTML, the table of contents is generated from "bookmarks" within the PDF file. The hyperlinked table of contents can appear either at the beginning of the HTML file or in a separate frame.

Micro Focus recommends that you configure the conversion so that the table of contents appears in a separate frame (the `pdfframe.ini` template demonstrates how to do this). Export uses absolute positioning when converting a PDF file, that is, the text appears in the exact position as in the original document. Table of contents entries do not contain absolute positioning information. Therefore, if the main document and the table of contents are generated in the same output file, the table of contents entries might overlap the body text in the document.

NOTE: When PDF bookmarks are converted to a table of contents in HTML, the generated links do not lead to the exact location of the destination marker, but jump to the page on which the destination marker exists. This is similar to the behavior of the Adobe Acrobat Reader.

Disable Bookmark Conversion

By default, Export converts PDF bookmarks to a table of contents in the HTML output. However, you can configure Export not to generate a table of contents based on the PDF bookmarks.

To specify that PDF bookmarks are not converted and included in the HTML output

1. Create an instance of the `ConfigOption` class. Set the `OptionType` argument to `CFG_SUPPRESSTOCPRINTIMAGE`, and the `OptionValue` argument to `1`.
2. Call the `setConfigOption` method and pass the `ConfigOption` object.
3. Call a convert method. See the Javadoc in the directory `install\javaapi\javadoc`, where `install` is the path name of the Export installation directory.

NOTE: A table of contents is not generated when a PDF file does *not* contain bookmarks, or when `CFG_SUPPRESSTOCPRINTIMAGE` is set.

Convert Invisible Text

PDF documents sometimes contain invisible text, which you can search in Adobe PDF Reader but cannot view in a web browser.

Toggle Invisible Text

You can add a JavaScript button to the upper right corner of the exported page, which you can click to toggle between invisible and regular text. When you turn on invisible text, the invisible text is displayed and the regular content is hidden; when you turn off invisible text, the invisible text is hidden.

Invisible text is hidden by default. The toggle button appears only if invisible text is detected in the PDF document.

To add an invisible text toggle button

- Set the `CFG_PDFINVISTEXTTOGGLE` field of the `HtmlExport` object. The parameter passed in is the label name for the toggle button.

Specify Opacity of Invisible Text

Invisible text often occurs in PDF documents when the PDF software processes rasterized images through optical character recognition and then inserts the text in the PDF. You might want to display both the invisible text as well as the rasterized image. To do so, you can set the invisible text *opacity* as determined by an integer from 0 to 100, where 0 hides the invisible text and 100 displays it fully.

Invisible text opacity is set to 0 by default.

To set invisible text opacity

- Set the `CFG_SETPDFINVISIBLETEXTOPACITY` field of the `HtmlExport` object to a value between 0 and 100.

Convert Rotated Text

By default, rotated text is displayed in its original position, at the original font size, and at 0 degrees rotation in the HTML output. The text is not rotated in the HTML output because text rotation is not supported by HTML.

Because the text is the original size, but might be displayed in a smaller space (at 0 degrees), the text might overlap adjacent text in the HTML output. To avoid this problem, you can specify that the rotated text be removed from its original position and displayed at the bottom of the HTML page on which it appears.

[[FUTURE: This option only applies to PDF. It will support other formats in future releases.]]

To specify that rotated text be displayed at the bottom of the HTML page

1. Create an instance of the `ConfigOption` class. Set the `OptionType` argument to `CFG_SETTEXTROTATE`, and the `OptionValue` argument to 1.
2. Call the `setConfigOption` method and pass the `ConfigOption` object.
3. Call a convert method. See the Javadoc in the directory `install\javaapi\javadoc`, where `install` is the path name of the Export installation directory.

NOTE: When this feature is enabled, white space is added to the bottom of every HTML page to accommodate any rotated text.

Control Hyphenation

There are two types of hyphens in a PDF document:

- A *soft hyphen* is added to a word by a word processor to divide the word across two lines. This is a discretionary hyphen and is used to ensure proper text flow in justified text.
- A *hard hyphen* is intentionally added to a word regardless of the word's position in the text flow. It is required by the rules of grammar, word usage or both. For example, compound words, such as "three-week vacation" and "self-confident" contain hard hyphens.

By default, KeyView maintains the source document's soft hyphens in the output HTML to more accurately represent the layout of the source document. However, if you are using Export to generate text output for an indexing engine, or if you are not concerned with maintaining the layout of the document, Micro Focus recommends that you remove soft hyphens from the HTML output. To remove soft hyphens, you must enable the soft hyphen flag.

NOTE: If the soft hyphen flag is enabled, every hyphen at the end of a line is considered a soft hyphen and removed from the HTML output. Hard hyphens at the end of a line are also

removed. This might result in an intentionally hyphenated word being extracted without a hyphen.

To remove soft hyphens from the HTML output

1. Create an instance of the `ConfigOption` class. Set the `OptionType` argument to `CFG_DELSOFTHYPHEN` and the `OptionValue` argument to `1`.
2. Call the `setConfigOption` method and pass the `ConfigOption` object.
3. Call a `convert` method. See the Javadoc in the directory `install\javaapi\javadoc`, where `install` is the path name of the Export installation directory.

Extract Custom Metadata from PDF Files

To extract custom metadata from your PDF files, add the custom metadata names to the `pdfsr.ini` file provided, and copy the modified file to the `\bin` directory. You can then extract metadata as you normally would.

The `pdfsr.ini` is in the `samples\pdfini` directory, and has the following structure:

```
<META>
<TOTAL>total_item_number</TOTAL>,
/metadata_tag_name datatype,
</META>
```

| Parameter | Description |
|-------------------|--|
| total item number | The total number of metadata tags that are listed. |
| metadata_tag_name | The metadata tag name used in the PDF files. |
| datatype | The data type of the metadata field. The possible types are: <ul style="list-style-type: none">• KV_String• KV_Int4• KV_DateTime• KV_ClipBoard• KV_Bool• KV_Unicode• KV_IEEE8• KV_Other |

For example:

```
<META>
<TOTAL> 4 </TOTAL>
/part_number      INT4
```

```
/volume          INT4  
/purchase_date   DATETIME  
/customer        STRING  
</META>
```

NOTE: Metadata cannot be extracted from PDFs when the PDF is converted to JPEG. See [Convert PDF Files to Raster Images, on page 87](#).

Convert Spreadsheet Files

Export has special configuration options that allow greater control over the conversion of spreadsheet files.

Convert Hidden Text in Microsoft Excel Files

Normally, Export does not convert hidden text from a Microsoft Excel spreadsheet because it is assumed that the text should not be exposed. You can change this default behavior and convert text in hidden rows, columns, and sheets by adding the following lines to the `formats_e.ini` file:

```
[Options]  
gethiddeninfo=1
```

Convert Headers and Footers in Microsoft Excel 2003 Files

Normally, Export does not convert headers and footers from Microsoft Excel 2003 spreadsheets. You can change this default behavior and convert headers and footers by adding the following lines to the `formats_e.ini` file:

```
[Options]  
ShowHeaderFooter=1
```

Specify Date and Time Format on UNIX Systems

System date and time format information is not stored in Microsoft Excel files. On Windows systems, you can specify a locale setting to determine the date and time format. However, on UNIX systems, the date and time format is set to the U.S. short date format by default (mm/dd/yyyy). To change the format, you must use a `formats_e.ini` option.

To specify the system date and time format on UNIX systems

- In the `formats_e.ini` file, specify the following options:

```
SysDateTime  
SysLongDate  
SysShortDate  
SysTime
```

NOTE:
These values cannot contain spaces.

For example, if you specify `SysDateTime=%d/%m/%Y`, dates and times are extracted in the following format:

`28/02/2008`

The format arguments are the same as those for the `strftime()` function. Refer to the following web page for more information.

<http://linux.die.net/man/3/strftime>

Convert Very Large Numbers in Spreadsheet Cells to Precision Numbers

You can now export numbers in Microsoft Excel files and write them to the output without formatting. By default, numbers are exported in the format specified by the Excel file (for example, *General*, *Currency*, and *Date*). Spreadsheets might contain cells that have very large numbers in them. Excel displays the numbers in a scientific notation that rounds or truncates the numbers.

To export numbers without formatting, add the following lines to the `formats_e.ini` file:

```
[Options]
ignoredefnumformats=1
```

Extract Microsoft Excel Formulas

Normally, the actual value of a formula is extracted from an Excel spreadsheet; the formula from which the value is derived is not included in the output. However, KeyView enables you to include the value as well as the formula in the output. For example, if Export is configured to extract the formula and the formula value, the output might look like this:

`245 = SUM(B21:B26)`

The calculated value from the cell is 245 and the formula from which the value is derived is `SUM(B21:B26)`.

NOTE: Depending on the complexity of the formulas, enabling formula extraction might result in slightly slower performance.

To set the extraction option for formulas, add the following lines to the `formats_e.ini` file:

```
[Options]
getformulastring=option
```

where *option* is one of the following.

| Option | Description |
|--------|--|
| 0 | Extract the formula value only. This is the default. |

| Option | Description |
|--------|--|
| | Set this option if formula extraction is enabled, and you want to return to the default. |
| 1 | Extract the formula only. |
| 2 | Extract the formula and the formula value. |

NOTE: If a function in a formula is not supported or is invalid, and option 1 or 2 is specified, only the calculated value is extracted. See the following table for a list of supported functions.

When formula extraction is enabled, Export can extract Microsoft Excel formulas that contain the functions listed in the following table:

Supported Microsoft Excel Functions

| | | | |
|--------------|-------------|------------|----------------|
| =ABS() | =ACOS() | =AND() | =AREAS() |
| =ASIN() | =ATAN2() | =ATAN2() | =AVERAGE() |
| =CELL() | =CHAR() | =CHOOSE() | =CLEAN() |
| =CODE() | =COLUMN() | =COLUMNS() | =CONCATENATE() |
| =COS() | =COUNT() | =COUNTA() | =DATE() |
| =DATEVALUE() | =DAVERAGE() | =DAY() | =DCOUNT() |
| =DDB() | =DMAX() | =DMIN() | =DOLLAR() |
| =DSTDEV() | =DSUM() | =DVAR() | =EXACT() |
| =EXP() | =FACT() | =FALSE() | =FIND() |
| =FIXED() | =FV() | =GROWTH() | =HLOOKUP() |
| =HOUR() | =ISBLANK() | =IF() | =INDEX() |
| =INDIRECT() | =INT() | =IPMT() | =IRR() |
| =ISERR() | =ISERROR() | =ISNA() | =ISNUMBER() |
| =ISREF() | =ISTEXT() | =LEFT() | =LEN() |
| =LINEST() | =LN() | =LOG() | =LOG10() |
| =LOGEST() | =LOOKUP() | =LOWER() | =MATCH() |
| =MAX() | =MDETERM() | =MID() | =MIN() |
| =MINUTE() | =MINVERSE() | =MIRR() | =MMULT() |
| =MOD() | =MONTH() | =N() | =NA() |
| =NOT() | =NOW() | =NPER() | =NPV() |
| =OFFSET() | =OR() | =PI() | =PMT() |

| | | | |
|--------------|------------|--------------|---------------|
| =PPMT() | =PRODUCT() | =PROPER() | =PV() |
| =RATE() | =REPLACE() | =REPT() | =RIGHT() |
| =ROUND() | =ROUND() | =ROW() | =ROWS() |
| =SEARCH() | =SECOND() | =SIGN() | =SIN() |
| =SLN() | =SQRT() | =STDEV() | =SUBSTITUTE() |
| =SUM() | =SYD() | =T() | =TAN() |
| =TEXT() | =TIME() | =TIMEVALUE() | =TODAY() |
| =TRANSPOSE() | =TREND() | =TRIM() | =TRUE() |
| =TYPE() | =UPPER() | =VALUE() | =VAR() |
| =VLOOKUP() | =WEEKDAY() | =YEAR() | |

Convert XML Files

Export enables you to extract all content or selected content from source XML files. It detects the following XML formats:

- generic XML
- Microsoft Office 2003 XML (Word, Excel, and Visio)
- StarOffice/OpenOffice XML (text document, presentation, and spreadsheet)

See [File Format Detection, on page 232](#) for more information on format detection.

Configure Element Extraction for XML Documents

When converting XML files, you can specify which elements and attributes are extracted according to the file's format ID or *root element*. This is useful when you want to extract only relevant text elements, such as abstracts from reports, or a list of authors from an anthology.

A root element is an element in which all other elements are contained. In the XML sample below, `book` is the root element:

```
<book>
  <title>XML Introduction</title>
  <product id="33-657" status="draft">XML Tutorial</product>
  <chapter>Introduction to XML
    <para>What is HTML</para>
    <para>What is XML</para>
  </chapter>
  <chapter>XML Syntax
    <para>Elements must have a closing tag</para>
    <para>Elements must be properly nested</para>
  </chapter>
</book>
```

For example, you could specify that when converting files with the root element `book`, the element `title` is extracted as metadata, and only `product` elements with a `status` attribute value of `draft` are extracted.

When you extract an element, the child elements within the element are also extracted. For example, if you extract the element `chapter` from the sample above, the child element `para` is also extracted.

Export defines default element extraction settings for the following XML formats:

- generic XML
- Microsoft Office 2003 XML (Word, Excel, and Visio)
- StarOffice/OpenOffice XML (text document, presentation, and spreadsheet)

These settings are defined internally and are used when converting these file formats; however, you can modify their values.

In addition to the default extraction settings, you can also add custom settings for your own XML document types. If you do not define custom settings for your own XML document types, the settings for the generic XML are used.

Modify Element Extraction Settings

You can modify configuration settings for XML documents through either the API or the `kvxconfig.ini` file.

NOTE: You can use customized element extraction settings only when converting files in process. When converting out of process, the default extraction settings are used.

Use the Java API

You can use the Java API to modify the settings for the standard XML document types, or to add configuration settings for your own XML document types.

To modify settings

1. Declare an array of `XMLConfigSet` objects.
2. Create an instance of the `ConfigOption` class with the following arguments:
 - a. Set the `OptionType` to `CFG_SETXMLCONFIGINFO`.
 - b. Set the `OptionValue` to `0`.
 - c. Set `OptionData` to the array object.
3. Call the `setConfigOption` method, and pass the `ConfigOption` object.
4. Call a convert method. For example:

```
XMLConfigSet[] XMLInfo;  
ConfigOption config=new ConfigOption(Export.CFG_SETXMLCONFIGINFO, 0, XMLInfo);  
objExport.setConfigOption(config);
```

Use an Initialization File

You can use the initialization file to modify the settings for the standard XML document types, or to add configuration settings for your own XML document types.

To modify settings

- 1. Modify the kvxconfig.ini file.
- 2. Use the template file when processing the XML file.

The Java sample program (HtmlConvFileToFile) demonstrates how to use a template file during the conversion process. See [HtmlConvFileToFile, on page 114](#).

Modify Element Extraction Settings in the kvxconfig.ini File

The kvxconfig.ini file contains default element extraction settings for supported XML formats. The file is in the install\OS\bin directory, where install is the path name of the Export installation directory and OS is the name of the operating system.

For example, the following entry defines extraction settings for the Microsoft Visio 2003 XML format:

```
[config3]
eKVFormat=MS_Visio_XML_Fmt
szRoot=
szInMetaElement=DocumentProperties
szExMetaElement=PreviewPicture
szInContentElement=Text
szExContentElement=
szInAttribute=
```

The following options are available.

| Configuration Option | Description |
|----------------------|--|
| eKVFormat | The format ID as detected by the KeyView detection module. This option determines the file type to which these extraction settings apply. See File Format Detection, on page 232 for more information on format ID values. If you are adding configuration settings for a custom XML document type, this option is not defined. |
| szRoot | The file's root element. When the format ID is not defined, the root element is used to determine the file type to which these settings apply. To further qualify the element, specify its namespace. See Specify an Element's Namespace and Attribute, on the next page . |
| szInMetaElement | The elements extracted from the file as metadata. All other elements are extracted as text. |

| Configuration Option | Description |
|----------------------|---|
| | Separate multiple entries with commas. To further qualify the element, specify its namespace, its attributes, or both. See Specify an Element's Namespace and Attribute, below . |
| szExMetaElement | <p>The child elements in the included metadata elements that are not extracted from the file as metadata. For example, the default extraction settings for the Visio XML format extract the <code>DocumentProperties</code> element as metadata. This element includes child elements such as <code>Title</code>, <code>Subject</code>, <code>Author</code>, <code>Description</code>, and so on. However, the child element <code>PreviewPicture</code> is defined in <code>szExMetaElement</code> because it is binary data and should not be extracted.</p> <p>You cannot exclude any metadata elements from the output for StarOffice files. All metadata is extracted regardless of this setting.</p> <p>Separate multiple entries with commas. To further qualify the element, specify its namespace, its attributes, or both. See Specify an Element's Namespace and Attribute, below.</p> |
| szInContentElement | <p>The elements extracted from the file as content text. Enter an asterisk (*) to extract all elements including child elements.</p> <p>Separate multiple entries with commas. To further qualify the element, specify its namespace, its attributes, or both. See Specify an Element's Namespace and Attribute, below.</p> |
| szExContentElement | <p>The child elements in the included content elements that are not extracted from the file as content text.</p> <p>Separate multiple entries with commas. To further qualify the element, specify its namespace, its attributes, or both. See Specify an Element's Namespace and Attribute, below.</p> |
| szInAttribute | <p>The attribute values extracted from the file. If attributes are not defined here, attribute values are not extracted.</p> <p>Enter the namespace (if used), element name, and attribute name in the following format:</p> <p><i>namespace:elementname@attributename</i></p> <p>For example:</p> <p><code>microfocus:division@name</code></p> <p>Separate multiple entries with commas.</p> |

Specify an Element's Namespace and Attribute

To further qualify an element, you can specify that the element exist in a certain namespace, that it contain a specific attribute, or both. To define the namespace and attribute of an element, enter the following:

ns_prefix:elemname@attribname=attribvalue

Attribute values that contain spaces must be enclosed in quotation marks.

For example, the following entry:

```
bg:language@id=xml
```

extracts a language element in the bg namespace that contains the id attribute name with the value of "xml". This entry extracts the following element from an XML file:

```
<bg:language id="xml">XML is a simple, flexible text format derived from  
SGML</bg:language>
```

but does not extract:

```
<bg:language id="sgml">SGML is a system for defining markup  
languages.</bg:language>
```

or

```
<adv:language id="xml">The namespace should be a Uniform Resource Identifier  
(URI).</adv:language>
```

Add Configuration Settings for Custom XML Document Types

You can define element extraction settings for custom XML document types by adding the settings to the kvxconfig.ini file. For example, for files containing the root element microfocusxml, you could add the following section to the end of the initialization file:

```
[config101]  
eKVFormat=  
szRoot=microfocusxml  
szInMetaElement=dc:title,dc:meta@title,dc:meta@name=title  
szExMetaElement=  
  
szInContentElement=microfocus:division@name=dev,microfocus:division@name=export,p@  
style="Heading 1"  
szExContentElement=  
szInAttribute=microfocus:division@name
```

The custom extraction settings must be preceded by a section heading named [config*N*], where *N* is an integer starting at 100 and increasing by 1 for each additional file type, as in [config100], [config101], [config102], and so on. The default extraction settings for the supported XML formats are numbered config0 to config99. Currently only 0 to 6 are used.

Because a custom XML document type is not recognized by the KeyView detection module, the format ID is not defined. The file type is identified by the file's root element only.

If a custom XML document type is not defined in the kvxconfig.ini file or by the setConfigOption method, the default extraction settings for a generic XML document are used.

Error Messages

When a KeyView exception is thrown, it might be caused by one of the following errors.

| Exception | Description |
|--------------------------------|---|
| KVERR_Success | Function completed successfully. |
| KVERR_DLLNotFound | A DLL or shared library was not found. |
| KVERR_OutOfCore | Memory allocation failure. |
| KVERR_processCancelled | Callback function returns FALSE. |
| KVERR_badInputStream | Invalid or corrupt input stream. |
| KVERR_badOutputType | Invalid output is requested. |
| KVERR_General | General error. |
| KVERR_FormatNotSupported | File format is not supported. |
| KVERR_PasswordProtected | File is encrypted or password-protected. KeyView only supports secure PST, NSF, and ZIP files. |
| KVERR_ADSNotFound | Adobe Document Server not found. This error is obsolete. |
| KVERR_AutoDetFail | Autodetect error. |
| KVERR_AutoDetNoFormat | Unable to detect file format. |
| KVERR_ReaderInitError | Error initializing the reader. |
| KVERR_NoReader | No reader available for this format. |
| KVERR_CreateOutputFileFailed | Unable to create output file. If the overwrite flag in <code>setOverWrite</code> is FALSE and a subfile has the same name as a file in the target path, this error is generated. |
| KVERR_CreateTempFileFailed | Unable to create temporary file. |
| KVERR_ErrorWritingToOutputFile | Error writing to output file. |
| KVERR_CreateProcessFailed | Error creating a child process. |
| KVERR_WaitForChildFailed | Wait for child process failed. |
| KVERR_ChildTimeOut | Child process hung/timed out. |
| KVERR_ArchiveFileNotFound | Attempt to extract nonexistent file. |
| KVERR_ArchiveFatalError | Fatal error processing an archive file. |

| Exception | Description |
|--|---|
| KVError_OpenStreamFailure = KVERR_ArchiveFatalError +1 | Failed to open a stream during out-of-process filtering. This is used by KeyView Filter. |
| KVError_InterfaceFunctionNotFound | An interface function was not found during out-of-process filtering. This is used by KeyView Filter. |
| KVError_InputFileNotFound | Could not find the input file during out-of-process filtering. This is used by KeyView Filter. |
| KVError_OpenOutputFileFailed | Could not open the output file during out-of-process filtering. This is used by KeyView Filter. |
| KVError_MemoryLeak | Memory leak occurred during out-of-process filtering. This is used by KeyView Filter. |
| KVError_MemoryOverwrite | Memory overwrite occurred during out-of-process filtering. This is used by KeyView Filter. |
| KVError_GPF | Exception occurred during out-of-process filtering. This is used by KeyView Filter. |
| KVError_OopCore | Memory dump was generated in a child process during out-of-process filtering. This is used by KeyView Filter. |
| KVError_KVoopLogFailed | Creation of out-of-process error log failed. This is used by KeyView Filter. |
| KVError_OverNestedFileLimit | The container file has more than the allowable number of child documents. One or more child documents were not converted. Currently, this is not used. |
| KVError_PSTAccessFailed | <p>The PST file could not be converted. This error might be returned when a call to <code>extOpenDocument</code> returns NULL for one of the following reasons:</p> <ul style="list-style-type: none"> • Microsoft Outlook client is not installed • Microsoft Outlook client is installed, but is not the default email client • Microsoft Outlook client is installed, but is not configured correctly • PST file is corrupt • PST file is read-only (PST files must allow read and write access) • MAPI call fails • The bit editions of Microsoft Outlook do not match the bit editions of the KeyView software. |

| Exception | Description |
|------------------------------------|--|
| | For example, if 32-bit KeyView is used, 32-bit Outlook must be installed. If 64-bit KeyView is used, 64-bit Outlook must be installed. |
| KVError_PasswordRequired | To open the file, credentials must be provided. This error might be returned when a call to <code>extOpenDocument</code> returns NULL. |
| KVError_InvalidArgs | The input argument or structure is invalid. This is generated by the File Extraction APIs. |
| KVError_OutputFileExists | A file with the same name already exists in the output directory. This error is generated when extracting a subfile from a container file with the <code>setOverWrite</code> flag set to FALSE, and a file by the same name already exists in the output directory. |
| KVError_ReaderUsageDenied | <p>The current license key does not enable the document reader required to convert the file. This error might be returned when a call to <code>extOpenDocument</code> returns NULL.</p> <p>Some document readers are considered advanced features and are licensed separately from the KeyView SDK (for example, the PST and MBX readers). Contact your Micro Focus sales representative to get an updated license key</p> |
| KVError_OopBadConfig | Information in the <code>kvxconfig.ini</code> file is incomplete and cannot be used to convert the XML file. |
| KVError_OopBrokenPipe | Data was not transferred between the parent and child processes during out-of-process filtering because either the parent or child failed. This is used by KeyView Filter. |
| KVError_OopPipeOEF | Data was not transferred between the parent and child processes during out-of-process filtering because the parent process was shutdown. This is used by KeyView Filter. |
| KVError_IPCTimeOut | Either the parent or child process is waiting for a reply or request during out-of-process filtering. This is used by KeyView Filter. |
| KVError_InvalidOopDriverSignature | A client sent a request to the File Extraction out-of-process server, but context driver does not exist on the server. This is used by KeyView Filter. |
| KVError_InvalidOopServiceSignature | <p>A client sent a request to a File Extraction out-of-process server that does not exist.</p> <p>If this error is generated on the call to <code>fpClose()</code>, it can be ignored. This is used by KeyView Filter.</p> |

Show Hidden Data

Microsoft Word, Excel, and PowerPoint documents contain hidden information, some of which is shown by default when exported and some of which is hidden by default. There are several options that allow you to determine which types of hidden data are shown.

Hidden Data in Microsoft Documents

You can show several types of hidden data from Microsoft Word, Excel, and PowerPoint documents, each of which has a corresponding parameter in the `Export` class, which you can set to change the default behavior. The following table lists each data type, its default behavior, and its corresponding configuration parameter.

Hidden data settings

| Hidden Data Type | Default Behavior | Configuration API Parameter |
|-----------------------|--------------------|-------------------------------|
| Microsoft Word | | |
| Comments ¹ | Shown ² | CFG_WP_NOCOMMENTS |
| Hidden text | Hidden | CFG_WP_SHOWHIDDENTEXT |
| Date field codes | Calculated date | CFG_WP_SHOWDATEFIELDPCODE |
| File name field codes | Document file name | CFG_WP_SHOWFILENAMEFIELDPCODE |
| Microsoft Excel | | |
| Hidden information | Hidden | CFG_SS_SHOWHIDDENINFOR |
| Comments | Hidden | CFG_SS_SHOWCOMMENTS |
| Formulas | Calculated value | CFG_SS_SHOWFORMULA |
| Microsoft PowerPoint | | |
| Hidden slides | Shown | CFG_PG_HIDEHIDDENSLIDE |
| Comments | Shown ³ | CFG_PG_HIDECOMMENT |

¹You can also toggle Word comment settings with a configuration parameter in the `formats_e.ini` file. See [Toggle Word Comment Settings in the formats_e.ini File, on the next page](#).

²Shown by default in Microsoft Word 97 to 2003 documents.

³Shown by default in Microsoft PowerPoint 97 to 2000 documents.

Hidden data settings, continued

| Hidden Data Type | Default Behavior | Configuration API Parameter |
|--------------------------|------------------|--|
| Comments slide | Hidden | CFG_PG_SHOWCOMMENTSSSLIDE ¹ |
| Slide notes ² | Hidden | CFG_PG_SHOWSLIDENOTES |

To toggle the display of hidden data

- Use the `setConfigOption` method of the `Export` object, and set the `config` parameter to one of the options listed in [Hidden data settings, on the previous page](#). Setting a hidden data parameter changes the *default* behavior.

For example:

```
objExport.setConfigOption(Export.CFG_WP_SHOWHIDDENTEXT);
```

In this case, the configuration parameter ensures that hidden text in Microsoft Word documents is shown (it is hidden by default).

Toggle Word Comment Settings in the `formats_e.ini` File

You can also control Microsoft Word 97 to 2003 comment settings through a parameter in the `formats_e.ini` file.

The `formats_e.ini` file is in the `install\OS\bin` directory, where `install` is the path name of the Export installation directory and `OS` is the name of the operating system.

To toggle comment output in `formats_e.ini`

1. Open the `formats_e.ini` file in a text editor.
2. Under `[Options]`, add the `WP_NOCOMMENTS` parameter and set it to `0` to show comments, or to `1` to hide comments. For example:

```
[Options]  
WP_NOCOMMENTS=1
```

NOTE: The `CFG_WP_NOCOMMENTS` configuration API flag overrides the setting in `formats_e.ini`.

Toggle PowerPoint Slide Note Settings in the `formats_e.ini` File

You can also control Microsoft PowerPoint slide note settings through a parameter in the `formats_e.ini` file.

¹This setting affects PowerPoint 2003 and 2007 only.

²You can also toggle PowerPoint slide note settings with a configuration parameter in the `formats_e.ini` file. See [Toggle PowerPoint Slide Note Settings in the `formats_e.ini` File, below](#).

The `formats_e.ini` file is in the `install\OS\bin` directory, where `install` is the path name of the Export installation directory and `OS` is the name of the operating system.

To toggle slide note output in `formats_e.ini`

1. Open the `formats_e.ini` file in a text editor.
2. Under `[Options]`, add the `ShowSlideNotes` parameter and set it to `1` to show slide notes, or to `0` to hide slide notes. For example:

```
[Options]
ShowSlideNotes=1
```

NOTE: The `KVCFG_PG_SHOWSLIDENOTES` configuration API flag overrides the setting in `formats_e.ini`.

Exclude Japanese Guide Text

This option prevents output of Japanese phonetic guide text when Microsoft Excel (`.xlsx`) files are processed.

To prevent output of Japanese phonetic guide text

- Set `NoPhoneticGuides` to `TRUE` in the `formats_e.ini` file:

```
[Options]
NoPhoneticGuides=TRUE
```

You can also enable this option programatically when filtering by passing `KVFLT_NOPHONETICGUIDES` to `fpFilterConfig`.

Source Code Identification

When KeyView auto-detects a file that contains source code, it can attempt to identify the programming language that it is written in.

NOTE:
Source code identification is a new, experimental feature in KeyView 12.0. It is available only on Windows 64-bit, Linux 64-bit, and OSX 64-bit platforms.

You can set source code identification to different levels.

| Option | Description |
|-----------------------------------|--|
| <code>KVSOURCECODE_OFF</code> | Do not enable source code identification. |
| <code>KVSOURCECODE_ENABLED</code> | Enable source code identification for the most common source code formats. |

| | |
|---------------------------|--|
| KVSOURCECODE_ EXTENDED | Enable source code identification for all supported source code formats. This option might lead to false positives in some cases (for example, a C++ file might get identified as a rarer format). |
|---------------------------|--|

For the complete list of source code formats supported for both options, see [Detected Formats, on page 153](#).

You can enable source code identification by setting the appropriate level in the `formats_e.ini` file. For example:

```
[Options]  
SourceCodeDetection=KVSOURCECODE_ENABLED
```

Chapter 5: Sample Programs

This section describes the Java sample programs provided with HTML Export.

| | |
|--|-----|
| • Introduction | 109 |
| • ExtractExport | 109 |
| • HtmlTest | 112 |
| • HtmlConvFileToFile | 114 |
| • HtmlConvStreamToStream | 116 |

Introduction

The Java sample programs demonstrate how to use the Java implementation of HTML Export. The sample code is intended to provide a starting point for your own applications or to be used for reference purposes.

The following Java sample programs are provided:

- [ExtractExport](#), below
- [HtmlTest](#), on page 112
- [HtmlConvFileToFile](#), on page 114
- [HtmlConvStreamToStream](#), on page 116

The source code for the programs is in the directory *install\javaapi\sample*.

In addition to the sample programs, the following support files are also provided:

Sample configuration files:

- *HtmlSampleConfig_nt.txt* (Windows)
- *HtmlSampleConfig_unix.txt* (UNIX)

Batch and C shell files that run each program:

- *HtmlConvFileToFile.bat* (.csh)
- *HtmlConvStreamToStream.bat* (.csh)

These files are used by the *HtmlConvFileToFile* and *HtmlConvStreamToStream* programs.

ExtractExport

This program demonstrates the File Extraction interface and basic functionality of the Export interface. The *HtmlTest* sample program demonstrates more advanced functionality of the Export interface. See [HtmlTest](#), on page 112.

The program is available in J# (*ExtractExport.exe*) for .NET and Java (*ExtractExport.java*).

The ExtractExport program demonstrates the following functionality:

- opens a document
- extracts subfiles from a document
- repeats subfile extraction until all subfiles are extracted
- sets conversion options through a template file
- converts the subfile (or subfiles) and main file to HTML or XML
- enables you to specify the command-line options listed in [Options for the ExtractExport Sample Program, below](#)

To run ExtractExport

1. For the Java version of the program, add the location of the javaapi\KeyView.jar file, the javaapi\sample directory, and the Export bin directory to the CLASSPATH environment variable. (Not required for .NET version of the program.)
2. For the Java version of the program, type the following:

```
java -Djava.library.path=bin_directory ExtractExport [options] bin_
directoryinifileinput_fileoutput_file
```

For the .NET version of the program (ExtractExport.exe), type:

```
ExtractExport [options] bin_directoryinifileinput_fileoutput_file
```

where:

- bin_directory is the path to the Export bin directory.
- options is one or more of the options listed in [Options for the ExtractExport Sample Program, below](#).
- inifile is the path and file name of a template file.
- input_file is the path and file name of the source file.
- output_file is the path and file name of the output file if the source file is not a container file.

Options for the ExtractExport Sample Program

| Option | Description |
|----------------------|---|
| -extonly | This option extracts the subfiles from a source file, but does not convert the files after extraction. |
| -extdir directory | This option sets the suggested directory to which the subfiles are extracted. |
| -ext-fbody | This option extracts the formatted version of the message body (HTML or RTF) from mail files when possible. |
| -xml | This option converts the files to XML. The default is HTML. To use this option, XML |

Options for the ExtractExport Sample Program, continued

| Option | Description |
|----------------------------|--|
| | Export must be installed. The .NET version of the sample program does not support this option. |
| -source-cs charset | This option sets the character set of the source file. charset is a character set defined as a constant in the Export class. See Code Character Sets, on page 210 . |
| -target-cs charset | This option sets the character set of the output file. charset is a character set defined as a constant in the Export class. See Code Character Sets, on page 210 . |
| -little- end | This option sets the byte order for Unicode text to little endian. |
| -is | This option sets the input as a stream. The default is file. |
| -os | This option sets the output as a stream. The default is file. |
| -open-user username | This option specifies the user name used to open a protected PST or NSF file. |
| -open-pass password | This option specifies the password used to open a protected PST or NSF file. |
| -open- idfile idfile | This option specifies the user ID file used to open a protected PST or NSF file. |
| -open- createroot | This option creates a root directory on which a hierarchy can be based. See Create a Root Node, on page 50 . |
| -ext-nodir | This option specifies that the subfile directory structure is not created. |
| -ext- noheader | This option excludes mail header information from extracted message body text file. See Exclude Metadata from the Extracted Text File, on page 57 . |
| -meta outfile | This option extracts default mail metadata and writes it to a file. See Extract Mail Metadata, on page 52 . |
| -oop | This option converts the files in a separate process. See Convert Files Out of Process, on page 24 . |
| -ip | This option runs file extraction in the same process as the calling application (in process). See Convert Files Out of Process, on page 24 . |

HtmlTest

This program converts an input document to an output document and allows you to specify options in the command line. The command-line options are listed in [Options for the HTMLTest Sample Program, below](#). This program demonstrates most of the methods available in the Java API. The program is available in J# (HtmlTest.exe) for .NET and Java (HtmlTest.java).

To run HtmlTest

1. For the Java version of the program, add the location of the javaapi\KeyView.jar file, the javaapi\sample directory, and the Export bin directory to the CLASSPATH environment variable.
2. For the Java version of the program, type the following:

```
java -Djava.library.path=bin_directory HtmlTest [options] bin_directoryinifile  
inputfileoutputfile
```

For the .NET version of the program (HtmlTestJ.exe), type:

```
HtmlTestJExport [options] bin_directoryinifileinput_fileoutput_file
```

where:

- bin_directory is the path to the Export bin directory.
- options is one or more of the options listed in [Options for the HTMLTest Sample Program, below](#).
- inifile is the full path and file name of a template file. See [Set Conversion Options by Using the Template Files, on page 32](#).
- inputfile is the path and file name of the source file.
- outputfile is the path and file name of the generated file.

If a path is not specified, the file is output to the current directory.

Options for the HTMLTest Sample Program

| Option | Description |
|---------|---|
| -is | This option sets the input as a stream. The default is file. |
| -os | This option sets the output as a stream. The default is file. |
| -oop | This option runs Export as a separate process. See Convert Files Out of Process, on page 24 . The .NET version of the sample program does not support this option. |
| -oopksa | This option keeps a Servant process active after the Export out-of-process session is terminated. If the Servant remains active, subsequent conversion requests are processed more quickly because the Servant is already prepared to |

Options for the HTMLTest Sample Program, continued

| Option | Description |
|--------------------------|---|
| | <p>receive data.</p> <p>The .NET version of the sample program does not support this option.</p> |
| -x xmlconfigfile | <p>This option converts an XML file using customized extraction settings defined in the <code>kvxconfig.ini</code> file. If you do not enter the full path to the configuration file, the program looks for the file in the current working directory (<i>install\OS\bin</i>, where <i>install</i> is the path name of the Export installation directory and <i>OS</i> is the name of the operating system). See Convert XML Files, on page 97.</p> |
| -z tempdirectory | <p>This option specifies a temporary directory in which temporary files generated by the conversion process are stored.</p> <p>On Windows systems, there is a 64 K size limit to the temporary directory. After the limit is reached, you must either create a new directory or delete the contents of the existing directory; otherwise, you might receive an error message.</p> |
| -docformat formatfile | <p>This option extracts the file format information and writes it to a file.</p> <p><code>formatfile</code> is the name of the file to which the format information is written.</p> |
| -summary summaryfile | <p>This option extracts the metadata and writes it to a file.</p> <p><code>summaryfile</code> is the name of the file to which the metadata is written. See Extract Metadata, on page 69.</p> |
| -list listfile | <p>This option displays a list of the files that are automatically generated during the conversion.</p> <p><code>listfile</code> is the name of the file to which the file list is written.</p> |
| -hifipdf | <p>This option specifies the type of reader used to convert PDF documents. In HTML Export, PDF documents are converted using either the graphic-based PDF reader, named <code>kppdfldr</code>, or the basic PDF reader, named <code>pdfsr</code>. See Convert PDF Files to Raster Images, on page 87. By default, the basic reader (<code>pdfsr</code>) is used to convert PDF documents.</p> |
| -suptocim | <p>If you set this option, bookmarks in a PDF file are not used to generate a table of contents in the HTML output. By default, the table of contents is generated from "bookmarks" within the PDF file. See Convert PDF Files, on page 85.</p> |
| -textrotate | <p>This option specifies that rotated text should be displayed at the bottom of a page on which it appears. By default, rotated text in a file is displayed in its original position, at the original font size, and at 0 degrees rotation. The text is not rotated in the HTML output because text rotation is not supported by HTML. See Convert Rotated Text, on page 92.</p> <p>Currently, this configuration option only applies to PDF files.</p> |
| -pdforder | <p>This option specifies that PDF files are output in a logical reading order. The</p> |

Options for the HTMLTest Sample Program, continued

| Option | Description |
|-----------------------|---|
| orderFlag | orderFlag parameter is one of the following: <ul style="list-style-type: none">ltr—left-to-right paragraph direction.rtl—right-to-left paragraph direction.auto—The PDF reader determines the paragraph direction (left-to-right or right-to-left) for each PDF page, and then sets the direction accordingly.raw—Unstructured paragraph flow. Convert PDF Files to Raster Images, on page 87. |
| -hl term term term | This option specifies the text string or strings that are found and highlighted in the HTML output. You can specify a maximum of three terms. See Search and Highlight Terms, on page 84. |
| -hlcs charset | This option specifies the character set of the highlighted search terms in the HTML output. |
| -hlig | This option specifies that the text search is case insensitive. This option can be used only when the target character set for the highlighted search term is KVCS_1252. |
| -rm | If you set this option, text and graphics that were deleted from a document with revision tracking enabled are converted, and revision tracking information is included in the HTML output. See Convert Revision Tracking Information, on page 84. |
| -dsh | This option specifies that soft hyphens in PDF files are deleted from the converted output. See Control Hyphenation, on page 92. |

HtmlConvFileToFile

This program converts an input file to an output file using Java API calls in HTML Export, and a template file. See [Set Conversion Options by Using the Template Files, on page 32](#) for more information on template files. The program is available in C# (HtmlConvFileToFile.cs) for .NET and Java (HtmlConvFileToFile.java).

This program demonstrates the following functions:

- Extracts file format information (document type, format, and version) if it is available in the source document.
- Extracts metadata if it is available in the source document. Although the program extracts all the metadata in the document, it displays only the first element of metadata.
- Displays a list of the files that are automatically generated during the conversion.

- Configures the PDF reader used in the conversion. See [Convert PDF Files to Raster Images, on page 87](#).

The basic PDF reader (`pdfsr`) is used by default in the sample program. To specify that the graphic-based PDF reader is to be used to convert PDF documents, remove the comment from the following line in the `HtmlConvFileToFile.java` file:

```
objHtmlExport.setConfigOption(new ConfigOption(HtmlExport.CFG_SETHIFIPDF, 1, null));
```

- Specifies the directory in which temporary files created during conversion processes are stored. To specify the temporary directory, remove the comment from the following line in the `HtmlConvFileToFile.java` file:

```
objHtmlExport.setConfigOption(new ConfigOption(Export.CFG_SETTEMPDIRECTIONS, 0, "C:\\tmp"));
```

NOTE: On Windows systems, there is a 64 K size limit to the temporary directory. After the limit is reached, you must either create a new directory or delete the contents of the existing directory; otherwise, you might receive an error message.

- Extracts elements from a source XML file based on the extraction settings in the `kvxconfig.ini` file. See [Convert XML Files, on page 97](#).

Run HtmlConvFileToFile on Windows

To run HtmlConvFileToFile on Windows

1. For the Java version of the program, set `INSTALL_DIR` to the Export SDK installation directory in the `HtmlConvFileToFile.bat` file.
2. In the first line of the `HtmlSampleConfig_nt.txt` file, specify the path to the template file used in the conversion. You can use the template files in the `install\htmlexport\programs\ini` directory, where `install` is the path name of the Export installation directory. If you are processing XML files, specify the location of the `kvxconfig.ini` file. This is in the `bin` directory.
3. In the second line of the `HtmlSampleConfig_nt.txt` file, specify the path to the Export SDK\bin directory.
4. Copy `HtmlSampleConfig_nt.txt` to `HtmlSampleConfig.txt`.
5. For the Java version of the program, run the batch file in the `install\javaapi\sample` directory. Type the following:

```
HtmlConvFileToFile inputfileoutputfile
```

For the .NET version of the program, compile the program in Visual Studio and run the program from the Export `bin` directory:

```
HtmlConvFileToFile.exe inputfileoutputfile
```

where:

`inputfile` is the path and file name of the source file.

`outputfile` is the path and file name of the generated HTML file.

Run HtmlConvFileToFile on UNIX

To run HtmlConvFileToFile on UNIX

1. In the `HtmlConvFileToFile.csh` file, set `MKENV` to the platform on which you are running, either `_hpux11`, `_ilnx21`, `_rs6k43`, or `_ssol26`.
2. In the `HtmlConvFileToFile.csh` file, set `INSTALL_DIR` to the Export installation directory.
3. In the first line of the `HtmlSampleConfig_unix.txt` file, specify the path to the template file used in the conversion. You can use the template files in the `install/htmlxpt/programs/unix/ini`, where `install` is the path name of the Export installation directory. If you are processing XML files, specify the location of the `kvxconfig.ini` file. This is stored in the `bin` directory.
4. In the second line of the `HtmlSampleConfig_unix.txt` file, specify the path to the `ExportSDK/bin` directory.

NOTE: If you have copied the `HtmlSampleConfig_unix.txt` and C shell file from a Windows system or edited the files on a Windows system, remove the carriage returns from these files. You can use a utility such as the ColdFusion function `stripcr()` (<http://www.macromedia.com>) to do this. This step is not required if the files are taken from the UNIX installation.

5. Copy `HtmlSampleConfig_unix.txt` to `HtmlSampleConfig.txt`.
6. Run the C shell file in the `install/javaapi/sample` directory. Type the following:

```
./HtmlConvFileToFile.csh inputfileoutputfile
```

where:

`inputfile` is the path and file name of the source file.

`outputfile` is the path and file name of the generated HTML file.

If a path is not specified, the HTML file is output to the current directory.

HtmlConvStreamToStream

This program converts an input stream to an output stream using Java API calls in HTML Export, and a template file. See [Set Conversion Options by Using the Template Files, on page 32](#) for more information on template files.

Run HtmlConvStreamToStream on Windows

To run HtmlConvStreamToStream on Windows

1. In the `HtmlConvStreamToStream.bat` file, set `INSTALL_DIR` to the Export installation directory.
2. In the first line of the `HtmlSampleConfig_nt.txt` file, specify the path to the template file used in the conversion. You can use the template files in the `install\htmllexport\programs\ini` directory, where `install` is the path name of the Export installation directory.
3. In the second line of the `HtmlSampleConfig_nt.txt` file, specify the path to the Export SDK\bin directory.
4. Copy `HtmlSampleConfig_nt.txt` to `HtmlSampleConfig.txt`.
5. Run the batch file in the `install\javaapi\sample` directory. Type the following:

```
HtmlConvStreamToStream inputfile
```

where:

`inputfile` is the path and file name of the source file.

The generated HTML is output to the current DOS prompt.

Run HtmlConvStreamToStream on UNIX

To run HtmlConvStreamToStream on UNIX

1. In the `HtmlConvStreamToStream.csh` file, set `MKENV` to the platform on which you are running, either `_hpux11`, `_ilnx21`, `_rs6k43`, or `_ssol26`.
2. In the `HtmlConvStreamToStream.csh` file, set `INSTALL_DIR` to the Export installation directory.
3. In the first line of the `HtmlSampleConfig_unix.txt` file, specify the path to the template file used in the conversion. You can use the template files in the `install/htmlxpt/programs/unix/ini` directory, where `install` is the path name of the Export installation directory.
4. In the second line of the `HtmlSampleConfig_unix.txt` file, specify the path to the ExportSDK/bin directory.

NOTE: If you have copied the `HtmlSampleConfig_unix.txt` and C shell file from a Windows system or edited the files on a Windows system, remove the carriage returns from these files. You can use a utility such as the ColdFusion function `stripcr()` (<http://www.macromedia.com>) to do this. This step is not required if the files are taken from the UNIX installation.

5. Copy `HtmlSampleConfig_unix.txt` to `HtmlSampleConfig.txt`.
6. Run the C shell file in the `install/javaapi/sample` directory. Type the following:

```
./HtmlConvStreamToStream.csh inputfile
```

where:

`inputfile` is the path and file name of the source file.

The generated HTML is output to the current console (standard out).

Part III: Appendixes

This section lists supported formats, supported character sets and redistributed files, and provides information on format detection.

- [Supported Formats](#)
- [Detected Formats](#)
- [Character Sets](#)
- [Extract and Format Lotus Notes Subfiles](#)
- [Export Tokens](#)
- [File Format Detection](#)
- [Files Required for Redistribution](#)
- [Password Protected Files](#)

Appendix A: Supported Formats

This section lists the file formats that KeyView can process (either filter, convert, or display).

- [Supported Formats](#) 121

Supported Formats

The tables in this section provide the following information:

- The file formats supported by the Filter API, Export API, Viewing API, and File Extraction API. The supported versions and the format's extension are also listed. All of the formats listed in this section can be detected by the KeyView format detection module (*kwad*). For a complete list of formats that can be detected, see [Detected Formats, on page 151](#).
- The file formats for which KeyView can detect and extract the character set and metadata information (properties such as title, author, and subject).

Even though a file format might be able to provide character set information, some documents might not contain character set information. Therefore, the document reader would not be able to determine the character set of the document. In this case, either the operating system code page or the character set specified in the API is used.

- The document reader used to filter each format.

Key to Support Tables

| Symbol | Description |
|--------|---|
| Y | The format is supported. You can extract metadata for this format. You can determine the character set for this format. |
| N | The format is not supported. You cannot extract metadata for this format. You cannot determine the character set for this format. |
| P | Partial metadata is extracted from this format. Some non-standard fields are not extracted. |
| T | Only text is extracted from this format. Formatting information is not extracted. |
| M | Only metadata (title, subject, author, and so on) is extracted from this format. Text and formatting information are not extracted. |

Archive Formats

Supported Archive Formats

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|--|---------|-----------------------------------|-----------|--------|--------|------|---------|----------|---------|---------------|
| 7-Zip | 4.57 | z7zsr, multiarcsr ¹ | 7Z | N | N | Y | Y | N | n/a | N |
| AD1 | n/a | ad1sr | AD1 | N | N | Y | Y | N | n/a | N |
| ARJ | n/a | multiarcsr | ARJ | N | N | N | Y | N | n/a | N |
| B1 | n/a | b1sr | B1 | N | N | Y | Y | N | n/a | N |
| BinHex | n/a | kvhqxsr | HQX | N | N | Y | Y | N | n/a | N |
| Bzip2 | n/a | bzip2sr | BZ2 | N | N | Y | Y | N | n/a | N |
| CPIO (copy-in-and-out archiver) | n/a | multiarcsr | | N | N | N | Y | N | n/a | N |
| Debian binary package | n/a | multiarcsr | DEB | N | N | N | Y | N | n/a | N |
| DOS/Windows Object Library | n/a | multiarcsr | LIB, A | N | N | N | Y | N | n/a | N |
| Expert Witness Compression Format (EnCase) | 6 | encasesr | E01, L01 | N | N | Y | Y | N | n/a | N |
| | 7 | encase2sr | Lx01 | N | N | Y | Y | N | n/a | N |

¹7zip is supported with the multiarcsr reader on some platforms for Extract.

Supported Archive Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|-----------------------------------|---------|------------|------------|--------|--------|------|---------|----------|---------|---------------|
| GZIP | 2 | kvgzsr | GZ | N | N | N | Y | N | n/a | N |
| | | kvgz | GZ | N | N | Y | N | N | n/a | N |
| ISO | n/a | isosr | ISO | N | N | Y | Y | N | n/a | N |
| Java Archive | n/a | unzip | JAR | N | N | Y | Y | N | n/a | N |
| Legato EMailXtender Archive | n/a | emxsr | EMX | N | N | Y | Y | N | n/a | N |
| LZMA compressed data | n/a | multiarcsr | LZMA | N | N | N | Y | N | n/a | N |
| MacBinary | n/a | macbinsr | BIN | N | N | Y | Y | N | n/a | N |
| Mac Disk Copy Disk Image | n/a | dmgsr | DMG | N | N | Y | Y | N | n/a | N |
| Mac OS-X (Mach-O) executable | n/a | multiarcsr | | N | N | N | Y | N | n/a | N |
| Microsoft Backup File | n/a | bkfsr | BKF | N | N | Y | Y | N | n/a | N |
| Microsoft Cabinet format | 1.3 | cabsr | CAB | N | N | Y | Y | N | n/a | N |
| Microsoft Compiled HTML Help | 3 | chmsr | CHM | N | N | Y | Y | N | n/a | N |
| Microsoft Compressed Folder | n/a | lzhsr | LZH LHA | N | N | N | Y | N | n/a | N |
| Microsoft Power BI Desktop format | n/a | unzip | PBIX | N | N | N | Y | N | n/a | N |
| MSI (Microsoft Installer) | n/a | multiarcsr | MSI | N | N | N | Y | N | n/a | N |

Supported Archive Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|-------------------------------------|-----------------|------------|-----------|--------|--------|------|---------|----------|---------|---------------|
| PKZIP | through 9.0 | unzip | ZIP | N | N | Y | Y | N | n/a | N |
| RAR archive | 2.0 through 3.5 | rarsr | RAR | N | N | N | Y | N | n/a | N |
| RAR5 archive | 5 | multiarcsr | RAR5 | N | N | N | Y | N | n/a | N |
| RPM (package manager file) | n/a | multiarcsr | RPM | N | N | N | Y | N | n/a | N |
| SUN PEX Binary Archive | n/a | multiarcsr | | N | N | Y | Y | N | n/a | N |
| Tableau Packaged Data Source format | n/a | unzip | TDSX | N | N | N | Y | N | n/a | N |
| Tableau Packaged Workbook format | n/a | unzip | TWBX | N | N | N | Y | N | n/a | N |
| Tape Archive | n/a | tarsr | TAR | N | N | Y | Y | N | n/a | N |
| UNIX Compress | n/a | kvzeesr | Z | N | N | N | Y | N | n/a | N |
| | | kvzee | Z | N | N | Y | N | N | n/a | N |
| UUEncoding | all versions | uudsr | UUE | N | N | Y | Y | N | n/a | N |
| XZ | n/a | multiarcsr | XZ | N | N | N | Y | N | n/a | N |
| Windows Imaging Format | n/a | multiarcsr | WIM | N | N | N | Y | N | n/a | N |

Supported Archive Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|--------------------------------|------------|------------|-----------|--------|--------|------|---------|----------|---------|---------------|
| Windows Scrap File | n/a | olesr | SHS | N | N | N | Y | N | n/a | N |
| WinZip | through 10 | unzip | ZIP | N | N | Y | Y | N | n/a | N |
| XAR (Extensible Archive) | n/a | multiarcsr | | N | N | N | Y | N | n/a | N |
| Zipped Keyhole Markup Language | n/a | unzip | ZIP | N | N | N | Y | N | n/a | N |

Binary Format

Supported Binary Formats

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|--------------|---------|--------|-----------|--------|--------|------|---------|----------|---------|---------------|
| Executable | n/a | exesr | EXE | N | N | Y | N | N | n/a | N |
| Link Library | n/a | exesr | DLL | N | N | Y | N | N | n/a | N |

Computer-Aided Design Formats

Supported CAD Formats

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|--------------------------|--|-----------------------------------|------------------|--------|--------|------|----------------|----------|---------|---------------|
| AutoCAD Drawing | R13, R14, R15/2000, 2004, 2007, 2010, 2013, 2018 | kpODArdr kpDWGrdr ¹ | DWG | Y | Y | Y | N | Y | Y | N |
| AutoCAD Drawing Exchange | R13, R14, R15/2000, 2004, 2007, 2010, 2013 | kpODArdr kpDXFrdr ² | DXF | Y | Y | Y | N | Y | Y | N |
| CATIA formats | 5 | kpCATrdr | CAT ³ | Y | N | N | N | Y | N | N |
| Microsoft Visio | 4, 5, 2000, 2002, 2003, 2007, 2010 ⁴ | vsdsr | VSD | Y | Y | Y | Y ⁵ | Y | Y | N |
| | | kpVSD2rdr | VSD, VSS VST | Y | Y | Y | N | Y | Y | N |

¹The kpODArdr reader can filter, export, and view all versions but is supported only on Windows, Linux, and OSX. The kpDWGrdr reader is used on AIX, FreeBSD, Solaris, and SPARC platforms, but does not support graphics for versions after 2004 or text for versions after 2013.

²The kpODArdr reader can filter, export, and view all versions but is supported only on Windows, Linux, and OSX. The kpDXFrdr reader is used on AIX, FreeBSD, Solaris, and SPARC platforms, but does not support graphics for versions after 2004.

³All CAT file extensions, for example CATDrawing, CATProduct, CATPart, and so on.

⁴Viewing and Export use the graphic reader, kpVSD2rdr for Microsoft Visio 2003, 2007, and 2010, and vsdsr for all earlier versions. Image fidelity in Viewing and Export is therefore only supported for versions 2003 and above. Filter uses the graphic reader kpVSD2rdr for Microsoft Visio 2003, 2007, and 2010, and vsdsr for all earlier versions.

⁵Extraction of embedded OLE objects is supported for Filter on Windows platforms only.

Supported CAD Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|---------------------|---------|--------------------|--|--------|--------|----------------|---------|----------|---------|---------------|
| | 2013 | ActiveX components | VSDM VSSM VSTM VSDX VSSX VSTX | N | N | Y ¹ | N | Y | N | N |
| | | kpVSDXrdr | VSDM VSSM VSTM VSDX VSSX VSTX | Y | Y | Y | Y | Y | Y | N |
| Unigraphics (UG) NX | | kpUGrdr | PRT | Y | N | N | N | N | N | N |

Database Formats

Supported Database Formats

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|----------------|----------|--------|-----------|--------|--------|------|---------|----------|---------|---------------|
| dBase Database | III+, IV | dbfsr | DBF | Y | Y | Y | N | N | N | N |

¹Visio 2013 is supported in Viewing only, with the support of ActiveX components from the Microsoft Visio 2013 Viewer. Image fidelity is supported but other features, such as highlighting, are not.

Supported Database Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|-------------------|--|--------|------------|--------|--------|------|---------|----------|----------------|---------------|
| Microsoft Access | 95, 97, 2000, 2002, 2003, 2007, 2010, 2013, 2016 | mdbsr | MDB, ACCDB | Y | T | T | N | N | Y ¹ | N |
| Microsoft Project | 2000, 2002, 2003, 2007, 2010, 2013, 2016 | mppsrs | MPP | Y | Y | Y | Y | Y | Y | N |

Desktop Publishing

Supported Desktop Publishing Formats

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|---------------------|------------|---------|-----------|--------|--------|------|---------|----------|---------|---------------|
| Microsoft Publisher | 98 to 2016 | mspubsr | PUB | Y | T | T | Y | Y | Y | N |

Display Formats

Supported Display Formats

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|-----------|------------|--------|-----------|--------|--------|------|----------------|----------|---------|---------------|
| Adobe PDF | 1.1 to 1.7 | pdfsr | PDF | Y | Y | N | Y ² | Y | Y | N |
| | | pdf2sr | PDF | N | Y | N | N | N | N | N |

¹Charset is not supported for Microsoft Access 95 or 97.

²Includes support for extraction of subfiles from PDF Portfolio documents.

Supported Display Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|--------|---------|------------------------|-----------|--------|--------|------|---------|----------|---------|---------------|
| | | kppdfrdr | PDF | N | Y | Y | N | N | N | N |
| | | kppdf2rdr ¹ | PDF | N | N | Y | N | N | N | N |

Graphic Formats

Supported Graphic Formats

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|-------------------------------------|-------------------------------|-----------------------|-----------|--------|--------|------|---------|----------|---------|---------------|
| Computer Graphics Metafile | n/a | kpcgmrdr ² | CGM | Y | Y | Y | N | N | N | N |
| CorelDRAW ³ | through 9.0 10, 11, 12, X3 | kpcdrdr | CDR | N | Y | Y | N | N | N | N |
| DCX Fax System | n/a | kpcxdr | DCX | N | Y | Y | N | N | N | N |
| Digital Imaging & Communications in | n/a | dcmsr | DCM | M | N | N | N | Y | N | N |

¹kppdf2rdr is an alternate graphic-based reader that produces high-fidelity output but does not support other features such as highlighting or text searching.

²Files with non-partitioned data are supported.

³CDR/CDR with TIFF header.

Supported Graphic Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|------------------------------------|-------------|-------------|-------------------------------|--------|--------|------|---------|----------|---------|---------------|
| Medicine (DICOM) | | | | | | | | | | |
| Encapsulated PostScript (raster) | TIFF header | kpepsrdr | EPS | N | Y | Y | N | N | N | N |
| Enhanced Metafile | n/a | kpemfrdr | EMF | Y | Y | Y | N | Y | N | N |
| GIF | 87, 89 | kpgifdrdr | GIF | N | Y | Y | N | N | N | N |
| | | gifsr | | M | M | N | N | Y | N | N |
| ISO-BMFF JPEG 2000 compound image | n/a | kpjp2000rdr | JPM | N | Y | Y | N | N | N | N |
| | | jp2000sr | | M | M | N | N | Y | N | N |
| ISO-BMFF JPEG 2000 image | n/a | kpjp2000rdr | JP2 | N | Y | Y | N | N | N | N |
| | | jp2000sr | | M | M | N | N | Y | N | N |
| ISO-BMFF JPEG 2000 with extensions | n/a | kpjp2000rdr | JPX | N | Y | Y | N | N | N | N |
| | | jp2000sr | | M | M | N | N | Y | N | N |
| JBIG2 | n/a | kpJBIG2rdr | JBIG2 | N | Y | Y | N | N | N | N |
| JPEG | n/a | kpjpgdrdr | JPEG | N | Y | Y | N | N | N | N |
| | | jpgsr | | M | M | N | N | Y | N | N |
| JPEG 2000 | n/a | kpjp2000rdr | JP2, JPF, J2K, JPWL, JPX, PGX | N | Y | Y | N | N | N | N |
| | | jp2000sr | | M | M | N | N | Y | N | N |

Supported Graphic Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|---|---------|-------------|------------|--------|--------|------|---------|----------|---------|---------------|
| JPEG 2000 PGX Verification Model image | n/a | kppj2000rdr | PGX | N | Y | Y | N | N | N | N |
| | | jp2000sr | | M | M | N | N | Y | N | N |
| Lotus AMIDraw Graphics | n/a | kpsdwrdr | SDW | N | Y | Y | N | N | N | N |
| Lotus Pic | n/a | kppicrdr | PIC | Y | Y | Y | N | N | N | N |
| Macintosh Raster | 2 | kppctrdr | PIC PCT | N | Y | Y | N | N | N | N |
| MacPaint | n/a | kpmacrdr | PNTG | N | Y | Y | N | N | N | N |
| Microsoft Office Drawing | n/a | kpmsohdr | MSO | N | Y | Y | N | N | N | N |
| Omni Graffiti | n/a | kpGFLrdr | GRAFFLE | Y | N | N | N | Y | Y | N |
| PC PaintBrush | 3 | kppcxrdr | PCX | N | Y | Y | N | N | N | N |
| Portable Network Graphics | n/a | kppngrdr | PNG | N | Y | Y | N | N | N | N |
| | | pngsr | PNG | M | M | N | N | Y | N | N |
| Scalable Vector Graphics | n/a | xmlsr | SVG | Y | T | T | N | Y | Y | N |
| SGI RGB Image | n/a | kpsgirdr | RGB | N | Y | Y | N | N | N | N |
| Sun Raster Image | n/a | kpsunrdr | RS | N | Y | Y | N | N | N | N |

Supported Graphic Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|-------------------------|--------------------------|----------|-----------|----------------|--------|------|---------|----------|---------|---------------|
| Tagged Image File | through 6.0 ¹ | tifsr | TIFF | M | M | N | N | Y | N | N |
| | | kptifdr | TIFF | N | Y | Y | N | N | N | N |
| Truevision Targa | 2 | kpTGArdr | TGA | N | Y | Y | N | N | N | N |
| Windows Animated Cursor | n/a | kpanirdr | ANI | N | Y | Y | N | N | N | N |
| Windows Bitmap | n/a | kpbmprdr | BMP | N | Y | Y | N | N | N | N |
| | | bmpsr | BMP | M | M | N | N | Y | N | N |
| Windows Icon Cursor | n/a | kpicordr | ICO | N | Y | Y | N | N | N | N |
| Windows Metafile | 3 | kpwmfrdr | WMF | Y ² | Y | Y | N | N | N | N |
| WordPerfect Graphics 1 | 1 | kpwpgrdr | WPG | N | Y | Y | N | N | N | N |
| WordPerfect Graphics 2 | 2, 7 | kpwg2rdr | WPG | N | Y | Y | N | N | N | N |

¹The following compression types are supported: no compression, CCITT Group 3 1-Dimensional Modified Huffman, CCITT Group 3 T4 1-Dimensional, CCITT Group 4 T6, LZW, JPEG (only Gray, RGB and CMYK color space are supported), and PackBits.

²Windows Metafiles can contain both raster images (KeyView file class 4) and vector graphics (KeyView file class 5). Filtering is supported only for vector graphics (class 5).

Mail Formats

Supported Mail Formats

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|-------------------------------------|--------------------------------|--------------------|-----------|--------|--------|------|---------|----------|---------|---------------|
| Documentum EMCMF | n/a | msgsr | EMCMF | N | N | Y | Y | Y | Y | N |
| Domino XML Language ¹ | n/a | dxlsr | DXL | N | N | Y | Y | Y | N | N |
| GroupWise FileSurf | n/a | gwfssr | GWFS | N | N | Y | Y | Y | N | N |
| Legato Extender | n/a | onmsr | ONM | N | N | Y | Y | Y | N | N |
| Lotus Notes database | 4, 5, 6.0, 6.5, 7.0, 8.0 | nsfsr | NSF | N | N | Y | Y | Y | N | N |
| Mailbox ² | Thunderbird 1.0, Eudora 6.2 | mbxsr ³ | MBX | N | N | T | Y | Y | Y | N |
| Microsoft | 2004 | entsr | various | N | N | Y | Y | Y | Y | N |

¹Supports non-encrypted embedded files only.

²KeyView supports MBX files created by Eudora Email and Mozilla Thunderbird. MBX files created by other common mail applications are typically filtered, converted, and displayed.

³This reader supports both clear signed and encrypted S/MIME. KeyView supports S/MIME for PST, EML, MBX, and MSG files.

Supported Mail Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|--|--|--------------------|-----------|--------|--------|------|---------|----------|----------------|---------------|
| Entourage Database | | | | | | | | | | |
| Microsoft Outlook | 97, 2000, 2002, 2003, 2007, 2010, 2013, 2016, 2019 | msgsr ¹ | MSG, OFT | Y | T | T | Y | Y | Y ² | N |
| Microsoft Outlook DBX | 5.0, 6.0 | dbxsr | DBX | N | N | Y | Y | Y | Y | N |
| Microsoft Outlook Express | Windows 6 MacIntosh 5 | emlsr ³ | EML | Y | T | T | Y | Y | Y | N |
| | | mbxsr ⁴ | EML | N | N | T | Y | Y | Y | N |
| Microsoft Outlook iCalendar | 1.0, 2.0 | icssr | ICS, VCS | N | N | Y | Y | Y | Y | N |
| Microsoft Outlook for Macintosh | 2011 | olmsr | OLM | N | N | Y | Y | N | Y | N |
| Microsoft Outlook Offline Storage File | 97, 2000, 2002, 2003, 2007, 2010, 2013 | pffsr ⁵ | OST | N | N | Y | Y | Y | Y | N |

¹This reader supports both clear signed and encrypted S/MIME. KeyView supports S/MIME for PST, EML, MBX, and MSG files.

²Returns "Unicode" character set for version 2003 and up, and "Unknown" character set for previous versions.

³This reader supports both clear signed and encrypted S/MIME. KeyView supports S/MIME for PST, EML, MBX, and MSG files.

⁴This reader supports both clear signed and encrypted S/MIME. KeyView supports S/MIME for PST, EML, MBX, and MSG files.

⁵The reader pffsr is available only on Windows and Linux.

Supported Mail Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|--|--|--------------------|-----------|--------|--------|------|---------|----------|---------|---------------|
| Microsoft Outlook Personal Folder ¹ | 97, 2000, 2002, 2003, 2007, 2010, 2013, 2016, 2019 | pstsr ² | PST | N | N | Y | Y | Y | N | N |
| | 97, 2000, 2002, 2003, 2007, 2010, 2013 | pstnsr | PST | N | N | Y | Y | Y | Y | N |
| | 97, 2000, 2002, 2003, 2007, 2010, 2013, 2016, 2019 | pstxsr | PST | N | N | Y | Y | Y | Y | N |
| Microsoft Outlook vCard Contact | 2.1, 3.0, 4.0 | vcfsr | VCF | Y | Y | T | N | Y | N | N |
| Text Mail (MIME) | n/a | emlsr ³ | various | Y | T | T | Y | Y | Y | N |
| | | mbxsr ⁴ | various | Y | T | T | Y | Y | Y | N |
| Transport Neutral Encapsulation Format | n/a | tnefsr | various | N | N | Y | Y | Y | Y | N |

¹KeyView provides several readers capable of processing PST files. The `pstsr` reader uses the Microsoft Messaging Application Programming Interface (MAPI), works only on Windows, and requires that you have Microsoft Outlook installed. The `pstxsr` reader is available for Windows (32-bit and 64-bit) and Linux (64-bit only) and does not require Microsoft Outlook. The `pstnsr` reader is an alternative reader that does not require Microsoft Outlook, for all platforms not supported by `pstxsr`. For more information about these readers, see "Extract Subfiles from Outlook Personal Folders Files" in Chapter 3.

²This reader supports both clear signed and encrypted S/MIME. KeyView supports S/MIME for PST, EML, MBX, and MSG files.

³This reader supports both clear signed and encrypted S/MIME. KeyView supports S/MIME for PST, EML, MBX, and MSG files.

⁴This reader supports both clear signed and encrypted S/MIME. KeyView supports S/MIME for PST, EML, MBX, and MSG files.

Multimedia Formats

Viewing SDK plays some multimedia files using the Windows Media Control Interface (MCI). MCI is a set of Windows APIs that communicate with multimedia devices.

Supported Multimedia Formats

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|--------------------------------------|---------|---------|-----------|--------|--------|------|---------|----------|---------|---------------|
| 3GPP video file | n/a | mpeg4sr | 3GP | M | N | N | N | Y | N | N |
| 3GPP2 video file | n/a | mpeg4sr | 3G2 | M | N | N | N | Y | N | N |
| Adobe Flash Player audio | n/a | mpeg4sr | F4A | M | N | N | N | Y | N | N |
| Adobe Flash Player audio book | n/a | mpeg4sr | F4B | M | N | N | N | Y | N | N |
| Adobe Flash Player protected video | n/a | mpeg4sr | F4P | M | N | N | N | Y | N | N |
| Adobe Flash Player video | n/a | mpeg4sr | F4V | M | N | N | N | Y | N | N |
| Apple ISO-BMFF QuickTime video | n/a | MCI | QT MOV | N | N | Y | N | N | N | N |
| Apple MPEG-4 Part 14 audio | n/a | mpeg4sr | M4A | M | N | N | N | Y | N | N |
| Apple MPEG-4 Part 14 audio book | n/a | mpeg4sr | M4B | M | N | N | N | Y | N | N |
| Apple MPEG-4 Part 14 protected audio | n/a | mpeg4sr | M4P | M | N | N | N | Y | N | N |
| Apple MPEG-4 Part 14 | n/a | mpeg4sr | M4V | M | N | N | N | Y | N | N |

Supported Multimedia Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|------------------------------------|---------------|-------------|-------------------|--------|--------|------|---------|----------|---------|---------------|
| video | | | | | | | | | | |
| Audible Enhanced Audiobook | n/a | mpeg4sr | AAX | M | N | N | N | Y | N | N |
| KDDI video file | n/a | MCI | | N | N | Y | N | N | N | N |
| Advanced Systems Format | 1.2 | asfsr | ASF WMA WMV | N | N | N | N | Y | N | N |
| Audio Interchange File Format | n/a | MCI | AIFF | N | N | Y | N | N | N | N |
| | | aiffsr | AIFF | M | N | N | N | Y | N | N |
| ISO-BMFF MPEG-4 with AVC extension | n/a | mpeg4sr | | M | N | N | N | Y | N | N |
| Microsoft Wave Sound | n/a | MCI | WAV | N | N | Y | N | N | N | N |
| | | riffr | WAV | M | N | N | N | Y | N | N |
| MIDI | n/a | MCI | MID | N | N | Y | N | N | N | N |
| Mobile QuickTime video | n/a | mpeg4sr | MQV | M | N | N | N | Y | N | N |
| Motion JPEG 2000 | n/a | kpjp2000rdr | MJ2 MJP2 | N | Y | Y | N | N | N | N |
| | | jp2000sr | | M | M | N | N | Y | N | N |
| MPEG-1 Audio layer 3 | ID3 v1 and v2 | MCI | MP3 | N | N | Y | N | N | N | N |
| | | mp3sr | MP3 | M | M | Y | N | Y | N | N |

Supported Multimedia Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|--|---------|---------|------------|--------|--------|------|---------|----------|---------|---------------|
| MPEG-1 Video | 2, 3 | MCI | MPG | N | N | Y | N | N | N | N |
| MPEG-2 Audio | n/a | MCI | MPEGA | N | N | Y | N | N | N | N |
| MPEG-21 | n/a | mpeg4sr | | M | N | N | N | Y | N | N |
| MPEG-4 Audio | n/a | mpeg4sr | MP4 3GP | M | N | N | N | Y | N | N |
| Nero AAC audio | n/a | mpeg4sr | | M | N | N | N | Y | N | N |
| Nero MPEG-4 profile | n/a | mpeg4sr | | M | N | N | N | Y | N | N |
| Nero MPEG-4 profile with AVC extension | n/a | mpeg4sr | | M | N | N | N | Y | N | N |
| NeXT/Sun Audio | n/a | MCI | AU | N | N | Y | N | N | N | N |
| NTT MPEG-4 | n/a | mpeg4sr | | M | N | N | N | Y | N | N |
| QuickTime Movie | 2, 3, 4 | MCI | QT MOV | N | N | Y | N | N | N | N |
| Sony PSP MPEG-4 | n/a | mpeg4sr | MP4 | M | N | N | N | Y | N | N |
| Sony XAVC video | n/a | mpeg4sr | | M | N | N | N | Y | N | N |
| Windows Video | 2.1 | MCI | AVI | N | N | Y | N | N | N | N |

NOTE:

Depending on the default multimedia player installed on your computer, the View API might not be able to play some supported multimedia formats. To play multimedia files, the View API uses the Windows Media Control Interface (MCI) to communicate with the multimedia player installed on your computer. If the player does not play a multimedia file that is supported by the Viewing SDK, the View API cannot

play the file.

If you cannot play a supported multimedia file by using the View API, install a different multimedia player or compressor/decompressor (codec) component.

Presentation Formats

Supported Presentation Formats

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|---------------------------------------|------------------------------|--------------------------|-------------|--------|--------|------|---------|----------|---------|---------------|
| Apple iWork Keynote | 2, 3, '08, '09 | kplWPGGrdr | GZ | Y | Y | Y | N | Y | Y | N |
| | '13, '16, '18 iCloud 2018 | kplWPG13rdr ¹ | KEY | Y | T | N | N | N | N | N |
| Applix Presents | 4.0, 4.2, 4.3, 4.4 | kpagrdr | AG | Y | Y | Y | N | N | N | N |
| Corel Presentations | 6, 7, 8, 9, 10, 11, 12, X3 | kpshwrdr | SHW | Y | Y | Y | N | N | N | N |
| Extensible Forms Description Language | n/a | kpXFDLrdr | XFD XFDL | Y | Y | Y | N | Y | Y | N |
| Lotus Freelance Graphics | 96, 97, 98, R9, 9.8 | kpprzrdr | PRZ | Y | Y | Y | N | N | N | N |
| Lotus Freelance Graphics 2 | 2 | kpprerdr | PRE | Y | Y | Y | N | N | N | N |

¹This reader is available only on Windows (32-bit and 64-bit), Linux (32-bit and 64-bit), and Solaris x86-64.

Supported Presentation Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|----------------------------------|------------------------------|----------|--|--------|--------|------|---------|----------|----------------|----------------|
| Macromedia Flash | through 8.0 | swfsr | SWF | Y | Y | Y | N | N | Y ¹ | N |
| Microsoft PowerPoint Macintosh | 98 | kpp40rdr | PPT | Y | Y | Y | N | N | N | N |
| | 2001, v.X, 2004 | kpp97rdr | PPT PPS POT | Y | Y | Y | N | P | Y | N |
| Microsoft PowerPoint PC | 4 | kpp40rdr | PPT | Y | Y | Y | N | P | N | N |
| Microsoft PowerPoint Windows | 95 | kpp95rdr | PPT | Y | Y | Y | N | P | Y | N |
| Microsoft PowerPoint Windows | 97, 2000, 2002, 2003 | kpp97rdr | PPT PPS POT | Y | Y | Y | Y | P | Y | Y ² |
| Microsoft PowerPoint Windows XML | 2007, 2010, 2013, 2016, 2019 | kpppxrdr | PPTX PPTM POTX POTM PPSX PPSM PPAM | Y | Y | Y | Y | Y | Y | Y |

¹The character set cannot be determined for versions 5.x and lower.

²Slide footers are supported for Microsoft PowerPoint 97 and 2003.

Supported Presentation Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|---|-------------------|----------|--------------------------|--------|--------|------|----------------|----------|---------|---------------|
| OASIS Open Document Format | 1, 2 ¹ | kpodfrdr | SXD SXI ODG ODP | Y | Y | Y | Y ² | Y | Y | N |
| OpenOffice Impress, LibreOffice Impress | 1 to 5 | sosr | SXI SXP ODP | Y | T | T | N | Y | Y | N |
| StarOffice Impress | 3, 4, 5 | kpsddrdr | SDA SDD | Y | T | N | N | N | N | N |
| | 6, 7, 8, 9 | sosr | SXI SXP ODP | Y | T | T | N | Y | Y | N |

¹Generated by OpenOffice Impress 2.0, StarOffice 8 Impress, and IBM Lotus Symphony Presentation 3.0.

²Supported using the olesr embedded objects reader.

Spreadsheet Formats

Supported Spreadsheet Formats

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|-------------------------|----------------------------|-----------------------|------------|--------|--------|------|---------|----------|---------|---------------|
| Apple iWork Numbers | '08, '09 | iwsssr | GZ | Y | Y | Y | N | Y | Y | N |
| | '13, '16, '18, iCloud 2018 | iwss13sr ¹ | NUMBERS | Y | T | T | N | N | Y | N |
| Applix Spreadsheets | 4.2, 4.3, 4.4 | assr | AS | Y | Y | Y | N | N | Y | N |
| Comma Separated Values | n/a | csvsr | CSV | Y | Y | Y | N | N | N | N |
| Corel Quattro Pro | 5, 6, 7, 8 | qpssr | WB2 WB3 | Y | Y | Y | N | P | Y | N |
| | X4 | qpwsr | QPW | Y | N | Y | N | P | Y | N |
| Data Interchange Format | n/a | difsr | | Y | Y | Y | N | N | N | N |
| Lotus 1-2-3 | 96, 97, R9, 9.8 | l123sr | 123 | Y | Y | Y | N | P | Y | N |
| Lotus 1-2-3 | 2, 3, 4, 5 | wkssr | WK4 | Y | Y | Y | N | N | Y | N |
| Lotus 1-2-3 Charts | 2, 3, 4, 5 | kpchtrdr | 123 | N | Y | Y | N | N | N | N |
| Microsoft Excel Charts | 2, 3, 4, 5, 6, 7 | kpchtrdr | XLS | N | Y | Y | N | N | N | N |

¹This reader is available only on Windows (32-bit and 64-bit), Linux (32-bit and 64-bit), and Solaris x86-64.

Supported Spreadsheet Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|--------------------------------------|---------------------------------|---------|--------------------------------------|--------|--------|------|----------------|----------|---------|---------------|
| Microsoft Excel Macintosh | 98, 2001, v.X, 2004 | xlssr | XLS | Y | Y | Y | Y ¹ | Y | Y | N |
| Microsoft Excel Windows | 2.2 through 2003 | xlssr | XLS XLW XLT XLA | Y | Y | Y | Y ² | Y | Y | Y |
| Microsoft Excel Windows XML | 2007, 2010, 2013, 2016, 2019 | xlxsxr | XLSX XLTX XLSM XLTM XLAM | Y | Y | Y | Y | Y | Y | Y |
| Microsoft Excel Binary Format | 2007, 2010, 2013, 2016 | xlsbsr | XLSB | Y | Y | Y | N | N | N | N |
| Microsoft Works Spreadsheet | 2, 3, 4 | mwssr | S30 S40 | Y | Y | Y | N | N | Y | N |
| OASIS Open Document Format | 1, 2 ³ | odfsssr | ODS SXC STC | Y | Y | Y | Y ⁴ | Y | Y | N |
| OpenOffice Calc, LibreOffice Calc | 1 to 5 | sosr | SXC ODS | Y | T | T | N | Y | Y | N |

¹Supported using the embedded objects reader `olesr`.

²Supported for versions 97 and higher using the embedded objects reader `olesr`.

³Generated by OpenOffice Calc 2.0, StarOffice 8 Calc, and IBM Lotus Symphony Spreadsheet 3.0.

⁴Supported using the embedded objects reader `olesr`.

Supported Spreadsheet Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|-----------------|------------|---------|------------|--------|--------|------|---------|----------|---------|---------------|
| | | | OTS | | | | | | | |
| StarOffice Calc | 3, 4, 5 | starcsr | SDC | Y | T | T | N | N | N | N |
| | 6, 7, 8, 9 | sosr | SXC ODS | Y | T | T | N | Y | Y | N |

Text and Markup Formats

Supported Text and Markup Formats

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|-----------------------------|------------------|--------|------------|--------|--------|------|---------|----------|---------|---------------|
| ANSI | n/a | afsr | TXT | Y | Y | Y | N | N | N | N |
| ASCII | n/a | afsr | TXT | Y | Y | Y | N | N | N | N |
| HTML | 3, 4 | htmsr | HTM | Y | Y | Y | N | P | Y | N |
| Microsoft Excel Windows XML | 2003 | xmlsr | XML | Y | T | T | N | Y | Y | N |
| Microsoft Word Windows XML | 2003 | xmlsr | XML | Y | T | T | N | Y | Y | N |
| Microsoft Visio XML | 2003 | xmlsr | VDX VTX | Y | T | T | N | Y | Y | N |
| MIME HTML | n/a | mhtsr | MHT | Y | Y | Y | N | Y | Y | N |
| Rich Text Format | 1 through 1.7 | rtfsr | RTF | Y | Y | Y | N | P | Y | Y |

Supported Text and Markup Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|---|---------|----------|-----------|--------|--------|------|---------|----------|---------|---------------|
| Tableau Data Source format | n/a | xmlsr | TDS | Y | T | T | N | Y | Y | N |
| Tableau Map Source format | n/a | xmlsr | TMS | Y | T | T | N | Y | Y | N |
| Tableau Preferences format | n/a | xmlsr | TPS | Y | T | T | N | Y | Y | N |
| Tableau Workbook format | n/a | xmlsr | TWB | Y | T | T | N | Y | Y | N |
| Unicode HTML | n/a | unihtmsr | HTM | Y | Y | Y | N | Y | Y | N |
| Unicode Text | 3, 4 | unisr | TXT | Y | Y | Y | N | N | Y | N |
| Vector Open Diagnostic Data Exchange Format | n/a | xmlsr | ODX | Y | T | T | N | Y | Y | N |
| XHTML | 1.0 | htmsr | HTM | Y | Y | Y | N | Y | Y | N |
| XML (generic) | 1.0 | xmlsr | XML | Y | T | T | N | Y | Y | N |

Word Processing Formats

Supported Word Processing Formats

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|-------------------------------------|-------------------------|---------|-----------|--------|--------|------|---------|----------|---------|---------------|
| Adobe FrameMaker Interchange Format | 5, 5.5, 6, 7 | mifsr | MIF | Y | Y | Y | N | N | Y | N |
| Apple iChat Log | 1, AV 2 AV 2.1, AV 3 | ichatsr | ICHAT | Y | Y | Y | N | N | N | N |

Supported Word Processing Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|-----------------------------------|--------------------------------|--------------------|-----------|--------|--------|------|---------|----------|---------|---------------|
| Apple iWork Pages | '08, '09 | iwwpsr | GZ | Y | Y | Y | N | Y | Y | N |
| | '13, '16, '18 iCloud 2018 | iwwp13sr 1 | PAGES | Y | T | T | N | N | N | N |
| Applix Words | 3.11, 4, 4.1, 4.2, 4.3, 4.4 | awsr | AW | Y | Y | Y | N | N | Y | Y |
| Corel WordPerfect Linux | 6.0, 8.1 | wp6sr | WPS | Y | Y | Y | N | P | Y | N |
| Corel WordPerfect Macintosh | 1.02, 2, 2.1, 2.2, 3, 3.1 | wpmsr | WPM | Y | Y | Y | N | N | Y | N |
| Corel WordPerfect Windows | 5, 5.1 | wosr | WO | Y | Y | Y | N | P | Y | Y |
| Corel WordPerfect Windows | 6, 7, 8, 9, 10, 11, 12, X3 | wp6sr | WPD | Y | Y | Y | N | P | Y | Y |
| DisplayWrite | 4 | dw4sr | IP | Y | Y | Y | N | N | Y | N |
| Folio Flat File | 3.1 | foliosr | FFF | Y | Y | Y | N | Y | Y | Y |
| Founder Chinese E- paper Basic | 3.2.1 | cebsr ² | CEB | Y | N | N | N | N | N | N |

¹This reader is available only on Windows (32-bit and 64-bit), Linux (32-bit and 64-bit), and Solaris x86-64.

²This reader is only supported on Windows 32-bit platforms.

Supported Word Processing Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|---|---------------------------|----------|----------------|--------|--------|------|---------|----------|---------|---------------|
| Fujitsu Oasys | 7 | oa2sr | OA2 | Y | Y | Y | N | P | N | N |
| Haansoft Hangul | 97 | hwpsr | HWP | Y | Y | Y | N | Y | Y | N |
| | 2002, 2005, 2007, 2010 | hwposr | HWP | Y | Y | Y | Y | Y | Y | N |
| Health level7 | 2.0 | hl7sr | HL7 | Y | Y | Y | N | Y | Y | N |
| IBM DCA/RFT (Revisable Form Text) | SC23-0758-1 | dcasr | DC | Y | Y | Y | N | N | Y | N |
| JustSystems Ichitaro | 8 to 2013, 2018 | jtdsr | JTD | Y | Y | Y | N | P | N | Y |
| Lotus AMI Pro | 2, 3 | lasr | SAM | Y | Y | Y | N | P | Y | Y |
| Lotus AMI Professional Write Plus | 2.1 | lasr | AMI | Y | Y | Y | N | N | N | Y |
| Lotus Word Pro | 96, 97, R9 | lwpsr | LWP | Y | Y | Y | N | P | N | Y |
| Lotus SmartMaster | 96, 97 | lwpsr | MWP | Y | Y | Y | N | N | N | N |
| Microsoft OneNote | 2007, 2010, 2013, 2016 | kpONErdr | ONE ONETOC2 | Y | Y | Y | Y | N | Y | N |
| Microsoft OneNote Alternate Format | 2007, 2010, 2013, 2016 | onealtsr | ONE ONETOC2 | Y | T | T | Y | N | N | N |

Supported Word Processing Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|---------------------------------|------------------------------|--------|------------------------------|--------|--------|------|----------------|----------|---------|---------------|
| Microsoft Word Macintosh | 4, 5, 6, 98 | mbsr | DOC | Y | Y | Y | N | Y | N | Y |
| | 2001, v.X, 2004 | mw8sr | DOC DOT | Y | Y | Y | Y ¹ | Y | Y | N |
| Microsoft Word PC | 4, 5, 5.5, 6 | mwsr | DOC | Y | Y | Y | N | N | N | Y |
| Microsoft Word Windows | 1.0, 2.0 | misr | DOC | Y | Y | Y | N | N | N | Y |
| Microsoft Word Windows | 6, 7, 8, 95 | mw6sr | DOC | Y | Y | Y | N | Y | Y | Y |
| Microsoft Word Windows | 97, 2000, 2002, 2003 | mw8sr | DOC DOT | Y | Y | Y | Y ² | Y | Y | Y |
| Microsoft Word Windows XML | 2007, 2010, 2013, 2016, 2019 | mwxsr | DOCX DOCX DOTX DOTM | Y | Y | Y | Y | Y | Y | Y |
| Microsoft Word Windows Flat XML | 2007, 2010, 2013, 2016 | mwxsr | XML | Y | Y | Y | Y | Y | Y | Y |
| Microsoft Works | 1, 2, 3, 4 | mswsr | WPS | Y | Y | Y | N | N | N | Y |

¹Supported using the embedded objects reader `olesr`.

²Supported using the embedded objects reader `olesr`.

Supported Word Processing Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|---------------------------------------|--------------------|---------|-------------------------|--------|--------|------|----------------|----------|---------|---------------|
| Microsoft Works | 6, 2000 | msw6sr | WPS | Y | Y | Y | N | N | N | Y |
| Microsoft Windows Write | 1, 2, 3 | mwsr | WRI | Y | Y | Y | N | N | Y | N |
| OASIS Open Document Format | 1, 2 ¹ | odfwpsr | ODT SXW STW | Y | Y | Y | Y ² | Y | Y | Y |
| Omni Outliner | v3, OPML, OOutline | oo3sr | OO3 OPML OOUTLINE | Y | Y | Y | N | N | Y | N |
| OpenOffice Writer, LibreOffice Writer | 1 to 5 | sosr | SXW ODT | Y | T | T | N | Y | Y | N |
| Open Publication Structure eBook | 2.0, 3.0 | epubsr | EPUB | Y | Y | Y | N | Y | Y | N |
| StarOffice Writer | 3, 4, 5 | starwsr | SDW | Y | T | T | N | N | N | N |
| | 6, 7, 8, 9 | sosr | SXW ODT | Y | T | T | N | Y | Y | N |
| Skype Log | 3 | skypesr | DBB | Y | Y | Y | N | N | N | N |
| WordPad | through 2003 | rtfsr | RTF | Y | Y | Y | N | P | Y | N |

¹Generated by OpenOffice Writer 2.0, StarOffice 8 Writer, and IBM Lotus Symphony Documents 3.0.

²Supported using the embedded objects reader `olesr`.

Supported Word Processing Formats, continued

| Format | Version | Reader | Extension | Filter | Export | View | Extract | Metadata | Charset | Header/Footer |
|--------------------------|---------|--------------------|-----------|--------|--------|------|---------|----------|---------|---------------|
| XML Paper Specification | n/a | xpssr | XPS | Y | T | T | N | N | N | N |
| XyWrite | 4.12 | xywsr | XY4 | Y | Y | Y | N | N | N | N |
| Yahoo! Instant Messenger | n/a | yimsr ¹ | DAT | Y | Y | Y | N | N | N | N |

¹To successfully use this reader, you must set the KV_YAHOO_ID environment variable to the Yahoo user ID. You can optionally set the KV_OTHER_YAHOO_ID environment variable to the other Yahoo user ID. If you do not set it, "Other" is used by default. If you enter incorrect values for the environment variables, erroneous data is generated.

Appendix B: Detected Formats

This section lists the file formats that KeyView can detect.

- [Key to Detected Formats Table](#) 151
- [Detected Formats](#) 153

Key to Detected Formats Table

The detected formats table includes the following information:

| Column | Description |
|-------------|--|
| Format Name | The format name that is returned by KeyView format detection. <ul style="list-style-type: none">• In the C API, these values are defined in the <code>ENdocFmt</code> enumeration in <code>adDocFmt.h</code>.• In the .NET API these values are defined in the <code>Autonomy.API.Export.DocFormat</code> enumeration.• In the Java API these values are defined in the <code>com.verity.api.DocFormat</code> enumeration. |
| Number | The format number that is returned by KeyView format detection. This is the value associated with the Format Name in the relevant enumeration. |
| Category | This value is used in the KeyView configuration file <code>formats.ini</code> to specify the reader to use to filter, export, or view the format. Several formats might have the same category value. |
| Description | A short description of the file format. |
| MIME Type | The MIME type (if any). |
| Extension | A list of common file extensions for the file format. <div>NOTE: This is not a complete list of file extensions. KeyView does not distinguish between file types based on their extension. Instead, it detects the file format based on the file content. This is more reliable because content cannot always be predicted from the file extension, and because some file extensions are associated with multiple formats.</div> |
| File Class | The KeyView file class. <ul style="list-style-type: none">• In the C API, these values are defined in the <code>ENdocClass</code> enumeration in <code>adinfo.h</code>.• In the .NET API these values are defined in the |

| | |
|--|---|
| | <p><code>Autonomy.API.Export.DocClass</code> enumeration.</p> <ul style="list-style-type: none">• In the Java API these values are defined in the <code>com.verity.api.DocClass</code> enumeration. |
|--|---|

Detected Formats

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|--------------------------|--------|----------|---|-----------------------|-----------|-----------------|
| Reserved_Fmt | -1 | -1 | | | | AutoDetNoFormat |
| Unknown_Fmt | 0 | 0 | | | | AutoDetNoFormat |
| AES_Multiplus_Comm_Fmt | 1 | 1 | Multiplus (AES) | | PTF | adWORDPROCESSOR |
| ASCII_Text_Fmt | 2 | 2 | Plain Text file | text/plain | TXT | adWORDPROCESSOR |
| MSDOS_Batch_File_Fmt | 3 | 2 | MS-DOS Batch File | application/x-bat | BAT | adEXECUTABLE |
| Applix_Alis_Fmt | 4 | 3 | APPLIX ASTERIX | | AX | adWORDPROCESSOR |
| BMP_Fmt | 5 | 4 | Windows Bitmap Image (BMP) | image/bmp | BMP | adRASTERIMAGE |
| CT_DEF_Fmt | 6 | 5 | Convergent Technologies DEF Comm. Format | | | adWORDPROCESSOR |
| Corel_Draw_Fmt | 7 | 6 | Corel Draw (up to version 13/X3) | application/coreldraw | CDR | adVECTORGRAPHIC |
| CGM_ClearText_Fmt | 8 | 8 | Computer Graphics Metafile (CGM) | | CGM | adVECTORGRAPHIC |
| CGM_Binary_Fmt | 9 | 8 | Computer Graphics Metafile (CGM) | image/cgm | CGM | adVECTORGRAPHIC |
| CGM_Character_Fmt | 10 | 8 | Computer Graphics Metafile (CGM) | | CGM | adVECTORGRAPHIC |
| Word_Connection_Fmt | 11 | 9 | Word Connection | | CN | adWORDPROCESSOR |
| COMET_TOP_Word_Fmt | 12 | 10 | Nixdorf COMET TOP Financial Accounting software | | | adWORDPROCESSOR |
| CEOWrite_Fmt | 13 | 11 | CEOWrite | | CW | adWORDPROCESSOR |
| DSA101_Fmt | 14 | 12 | DSA101 (Honeywell Bull) | | | adWORDPROCESSOR |
| DCA_RFT_Fmt | 15 | 13 | DCA-RFT (IBM Revisable Form) | application/dca-rft | RFT, DC | adWORDPROCESSOR |
| CDA_DDIF_Fmt | 16 | 14 | CDA / DDIF | | DDIF | adWORDPROCESSOR |
| DG_CDS_Fmt | 17 | 16 | DG Common Data Stream (CDS) | | CDS | adWORDPROCESSOR |
| Micrografx_Draw_Fmt | 18 | 18 | Windows Draw (Micrografx) | | DRW | adVECTORGRAPHIC |
| Data_Point_VistaWord_Fmt | 19 | 19 | Vistaword | | DV | adWORDPROCESSOR |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|--------------------------|--------|----------|--|----------------------------|-----------|-----------------|
| DECdx_Fmt | 20 | 20 | DECdx | | DX | adWORDPROCESSOR |
| Enable_WP_Fmt | 21 | 21 | Enable Word Processing | | WPF | adWORDPROCESSOR |
| EPSF_Fmt | 22 | 22 | Encapsulated PostScript | application/postscript | EPS | AutoDetNoFormat |
| Preview_EPSF_Fmt | 23 | 22 | Encapsulated PostScript | application/postscript | | AutoDetNoFormat |
| MS_Executable_Fmt | 24 | 23 | MSDOS/Windows Program | application/x-msdownload | EXE | adEXECUTABLE |
| G31D_Fmt | 25 | 24 | CCITT G3 1D | | | adRASTERIMAGE |
| GIF_87a_Fmt | 26 | 25 | Graphics Interchange Format (GIF87a) | image/gif | GIF | adRASTERIMAGE |
| GIF_89a_Fmt | 27 | 25 | Graphics Interchange Format (GIF89a) | image/gif | GIF | adRASTERIMAGE |
| HP_Word_PC_Fmt | 28 | 26 | HP Word PC | | HW | adWORDPROCESSOR |
| IBM_1403_LinePrinter_Fmt | 29 | 27 | IBM 1403 Line Printer | | I4 | adWORDPROCESSOR |
| IBM_DCF_Script_Fmt | 30 | 28 | DCF Script | | IC | adWORDPROCESSOR |
| IBM_DCA_FFT_Fmt | 31 | 29 | DCA-FFT (IBM Final Form) | | IF, FFT | adWORDPROCESSOR |
| Interleaf_Fmt | 32 | 30 | Interleaf | | | adWORDPROCESSOR |
| GEM_Image_Fmt | 33 | 31 | GEM Bit Image | | IMG | adRASTERIMAGE |
| IBM_Display_Write_Fmt | 34 | 32 | Display Write | | IP | adWORDPROCESSOR |
| Sun_Raster_Fmt | 35 | 33 | Sun Raster | image/x-cmu-raster | RAS | adRASTERIMAGE |
| Ami_Pro_Fmt | 36 | 35 | Lotus Ami Pro | application/x-lotus-amipro | SAM | adWORDPROCESSOR |
| Ami_Pro_StyleSheet_Fmt | 37 | 35 | Lotus Ami Pro Style Sheet | | | adWORDPROCESSOR |
| MORE_Fmt | 38 | 36 | MORE Database MAC | | | adOUTLINE |
| Lyrix_Fmt | 39 | 37 | Lyrix Word Processing | | | adWORDPROCESSOR |
| MASS_11_Fmt | 40 | 38 | MASS-11 | | M1 | adWORDPROCESSOR |
| MacPaint_Fmt | 41 | 39 | MacPaint | | PNTG | adRASTERIMAGE |
| MS_Word_Mac_Fmt | 42 | 40 | Microsoft Word for Macintosh (up to version 3) | application/msword | DOC | adWORDPROCESSOR |
| SmartWare_II_Comm_Fmt | 43 | 41 | SmartWare II | | | adCOMMUNICATION |
| MS_Word_Win_Fmt | 44 | 42 | Microsoft Word for Windows (up to version 6) | application/msword | DOC, WPS | adWORDPROCESSOR |
| Multimate_Fmt | 45 | 43 | MultiMate | | MM | adWORDPROCESSOR |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|----------------------------|--------|----------|--|-------------------------|-----------|-----------------|
| Multimate_Fnote_Fmt | 46 | 43 | MultiMate Footnote File | | | adWORDPROCESSOR |
| Multimate_Adv_Fmt | 47 | 43 | MultiMate Advantage | | | adWORDPROCESSOR |
| Multimate_Adv_Fnote_Fmt | 48 | 43 | MultiMate Advantage Footnote File | | | adWORDPROCESSOR |
| Multimate_Adv_II_Fmt | 49 | 43 | MultiMate Advantage II | | | adWORDPROCESSOR |
| Multimate_Adv_II_Fnote_Fmt | 50 | 43 | MultiMate Advantage II Footnote File | | FBX, FNX | adWORDPROCESSOR |
| Multiplan_PC_Fmt | 51 | 44 | Multiplan (PC) | | | adSPREADSHEET |
| Multiplan_Mac_Fmt | 52 | 44 | Multiplan (Mac) | | | adSPREADSHEET |
| MS_RTF_Fmt | 53 | 45 | Rich Text Format (RTF) | application/rtf | RTF | adWORDPROCESSOR |
| MS_Word_PC_Fmt | 54 | 46 | Microsoft Word for PC (up to version 6) | application/x-ms-wordpc | MW | adWORDPROCESSOR |
| MS_Word_PC_StyleSheet_Fmt | 55 | 46 | Microsoft Word for PC (up to version 6) Style Sheet | | | adWORDPROCESSOR |
| MS_Word_PC_Glossary_Fmt | 56 | 46 | Microsoft Word for PC (up to version 6) Glossary | | | adWORDPROCESSOR |
| MS_Word_PC_Driver_Fmt | 57 | 46 | Microsoft Word for PC (up to version 6) Driver | | | adWORDPROCESSOR |
| MS_Word_PC_Misc_Fmt | 58 | 46 | Microsoft Word for PC (up to version 6) Miscellaneous File | | | adWORDPROCESSOR |
| NBI_Async_Archive_Fmt | 59 | 47 | NBI Async Archive Format | | | adWORDPROCESSOR |
| Navy_DIF_Fmt | 60 | 48 | Navy DIF (document interchange format) | | ND | adWORDPROCESSOR |
| NBI_Net_Archive_Fmt | 61 | 49 | NBI Net Archive Format | | NN | adWORDPROCESSOR |
| NIOS_TOP_Fmt | 62 | 50 | NIOS TOP | | | adWORDPROCESSOR |
| FileMaker_Mac_Fmt | 63 | 51 | Filemaker MAC | | FP5, FP7 | adDATABASE |
| ODA_Q1_11_Fmt | 64 | 52 | ODA / ODIF Q1 11 | | OD | adWORDPROCESSOR |
| ODA_Q1_12_Fmt | 65 | 52 | ODA / ODIF Q1 12 | | OD | adWORDPROCESSOR |
| OLIDIF_Fmt | 66 | 53 | OLIDIF (Olivetti) | | | adWORDPROCESSOR |
| Office_Writer_Fmt | 67 | 55 | Office Writer | | OW | adWORDPROCESSOR |
| PC_Paintbrush_Fmt | 68 | 56 | PC Paintbrush Graphics (PCX) | image/vnd.zbrush.pcx | PCX | adRASTERIMAGE |
| CPT_Comm_Fmt | 69 | 57 | CPT Corporation word processor | | PF | adWORDPROCESSOR |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|----------------------------|--------|----------|--------------------------------------|---------------------------------|-----------|-----------------|
| Lotus_PIC_Fmt | 70 | 58 | Lotus PIC | image/x-pict | PIC | adVECTORGRAPHIC |
| Mac_PICT_Fmt | 71 | 59 | QuickDraw Picture | image/x-pict | PCT | AutoDetNoFormat |
| Philips_Script_Word_Fmt | 72 | 60 | Philips Script | | | adWORDPROCESSOR |
| PostScript_Fmt | 73 | 61 | PostScript | application/postscript | PS | adVECTORGRAPHIC |
| PRIMEWORD_Fmt | 74 | 62 | PRIMEWORD | | | adWORDPROCESSOR |
| Quadratron_Q_One_v1_Fmt | 75 | 63 | Q-One V1.93J | | Q1, QX | adWORDPROCESSOR |
| Quadratron_Q_One_v2_Fmt | 76 | 64 | Q-One V2.0 | | Q1, QX | adWORDPROCESSOR |
| SAMNA_Word_IV_Fmt | 77 | 65 | SAMNA Word | | SAM | adWORDPROCESSOR |
| Ami_Pro_Draw_Fmt | 78 | 66 | Lotus Ami Pro Draw | | SDW | adVECTORGRAPHIC |
| SYLK_Spreadsheet_Fmt | 79 | 67 | SYmbolic LinK (SYLK) format | | SLK | adSPREADSHEET |
| SmartWare_II_WP_Fmt | 80 | 68 | Informix SmartWare II word processor | | DOC | adWORDPROCESSOR |
| Symphony_Fmt | 81 | 69 | Lotus Symphony spreadsheet | | WR1 | adSPREADSHEET |
| Targa_Fmt | 82 | 70 | Targa image | image/x-tga | TGA | adRASTERIMAGE |
| TIFF_Fmt | 83 | 71 | Tag Image File Format (TIFF) | image/tiff | TIF, TIFF | adRASTERIMAGE |
| Targon_Word_Fmt | 84 | 72 | Targon Word | | TW | adWORDPROCESSOR |
| Uniplex_Ucalc_Fmt | 85 | 73 | Uniplex Ucalc | | SS | adSPREADSHEET |
| Uniplex_WP_Fmt | 86 | 74 | Uniplex word processor | | UP | adWORDPROCESSOR |
| MS_Word_UNIX_Fmt | 87 | 75 | Microsoft Word UNIX | application/msword | | adWORDPROCESSOR |
| WANG_PC_Fmt | 88 | 76 | WANG PC | | | adWORDPROCESSOR |
| WordERA_Fmt | 89 | 77 | WordERA | | | adWORDPROCESSOR |
| WANG_WPS_Comm_Fmt | 90 | 78 | WANG WPS | | WF | adWORDPROCESSOR |
| WordPerfect_Mac_Fmt | 91 | 79 | WordPerfect MAC | application/x-corel-wordperfect | | adWORDPROCESSOR |
| WordPerfect_Fmt | 92 | 86 | WordPerfect version 4 | application/x-corel-wordperfect | WP, WP4 | adWORDPROCESSOR |
| WordPerfect_VAX_Fmt | 93 | 139 | WordPerfect VAX | application/x-corel-wordperfect | | adWORDPROCESSOR |
| WordPerfect_Macro_Fmt | 94 | 139 | WordPerfect Macro | application/vnd.wordperfect | MRS | adWORDPROCESSOR |
| WordPerfect_Dictionary_Fmt | 95 | 139 | WordPerfect Spelling Dictionary | application/vnd.wordperfect | SPW | adWORDPROCESSOR |
| WordPerfect_Thesaurus_Fmt | 96 | 139 | WordPerfect Thesaurus | application/vnd.wordperfect | | adWORDPROCESSOR |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|-----------------------------|--------|----------|---|------------------------------------|-----------|------------------|
| WordPerfect_Resource_Fmt | 97 | 139 | WordPerfect Resource File | application/vnd.wordperfect | WWK, PRS | adWORDPROCESSOR |
| WordPerfect_Driver_Fmt | 98 | 139 | WordPerfect Driver | application/vnd.wordperfect | IRS, VRS | adWORDPROCESSOR |
| WordPerfect_Cfg_Fmt | 99 | 139 | WordPerfect Configuration File | application/vnd.wordperfect | PFX | adWORDPROCESSOR |
| WordPerfect_Hyphenation_Fmt | 100 | 139 | WordPerfect Hyphenation Dictionary | application/vnd.wordperfect | HYC | adWORDPROCESSOR |
| WordPerfect_Misc_Fmt | 101 | 139 | WordPerfect Miscellaneous File | application/vnd.wordperfect | | adWORDPROCESSOR |
| WordMARC_Fmt | 102 | 82 | WordMARC Composer | video/x-ms-wm | WM, PW | adWORDPROCESSOR |
| Windows_Metafile_Fmt | 103 | 83 | Windows Metafile | image/wmf | WMF | adVECTORGRAPHIC |
| Windows_Metafile_NoHdr_Fmt | 104 | 83 | Windows Metafile (no header) | image/wmf | WMF | adVECTORGRAPHIC |
| SmartWare_II_DB_Fmt | 105 | 84 | Informix SmartWare II database | | | adDATABASE |
| WordPerfect_Graphics_Fmt | 106 | 195 | WordPerfect Graphics (version 2 and higher) | application/vnd.wordperfect | WPG, QPG | AutoDetNoFormat |
| WordStar_Fmt | 107 | 87 | WordStar | | WS, WSD | adWORDPROCESSOR |
| WANG_WITA_Fmt | 108 | 88 | WANG WITA | | WT | adWORDPROCESSOR |
| Xerox_860_Comm_Fmt | 109 | 89 | Xerox 860 | | | adWORDPROCESSOR |
| Xerox_Writer_Fmt | 110 | 91 | Xerox Writer | | | adWORDPROCESSOR |
| DIF_SpreadSheet_Fmt | 111 | 92 | Data Interchange Format (DIF) | application/dif+xml | DIF | adSPREADSHEET |
| Enable_Spreadsheet_Fmt | 112 | 93 | Enable Spreadsheet | application/vnd.epson.ssf | SSF | adSPREADSHEET |
| SuperCalc_Fmt | 113 | 94 | Sorcim SuperCalc spreadsheet | | CAL | adSPREADSHEET |
| UltraCalc_Fmt | 114 | 95 | UltraCalc spreadsheet | | | adSPREADSHEET |
| SmartWare_II_SS_Fmt | 115 | 96 | Informix SmartWare II spreadsheet | | | adSPREADSHEET |
| SOF_Encapsulation_Fmt | 116 | 97 | Serialized Object Format (SOF) | application/java-serialized-object | SOF | adENCAPSULATION |
| PowerPoint_Win_Fmt | 117 | 98 | Microsoft PowerPoint PC (up to version 4) | application/x-ms-powerpoint | PPT | adPRESENTATION |
| PowerPoint_Mac_Fmt | 118 | 99 | Microsoft PowerPoint MAC (up to version 4) | application/x-ms-powerpoint | PPT | adPRESENTATION |
| PowerPoint_95_Fmt | 119 | 212 | Microsoft PowerPoint 95 | application/x-ms-powerpoint | PPT | adPRESENTATION |
| PowerPoint_97_Fmt | 120 | 272 | Microsoft PowerPoint 97 | application/x-ms-powerpoint | PPT | adPRESENTATION |
| PageMaker_Mac_Fmt | 121 | 100 | PageMaker for Macintosh | | | adDESKTOPPUBLISH |
| PageMaker_Win_Fmt | 122 | 101 | PageMaker for Windows | | | adDESKTOPPUBLISH |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|---------------------------|--------|----------|--|----------------------------|-----------|-----------------|
| MS_Works_Mac_WP_Fmt | 123 | 103 | Microsoft Works Word Processor for MAC | application/x-msworks | MWK | adWORDPROCESSOR |
| MS_Works_Mac_DB_Fmt | 124 | 104 | Microsoft Works Database for MAC | application/x-msworks | | adDATABASE |
| MS_Works_Mac_SS_Fmt | 125 | 105 | Microsoft Works Spreadsheet for MAC | application/x-msworks | | adSPREADSHEET |
| MS_Works_Mac_Comm_Fmt | 126 | 106 | Microsoft Works Communication for MAC | application/x-msworks | | adCOMMUNICATION |
| MS_Works_DOS_WP_Fmt | 127 | 107 | Microsoft Works Word Processor for DOS | application/x-msworks | WPS | adWORDPROCESSOR |
| MS_Works_DOS_DB_Fmt | 128 | 108 | Microsoft Works Database for DOS | application/x-msworks | WDB | adDATABASE |
| MS_Works_DOS_SS_Fmt | 129 | 109 | Microsoft Works Spreadsheet for DOS | application/x-msworks | | adSPREADSHEET |
| MS_Works_Win_WP_Fmt | 130 | 227 | Microsoft Works Word Processor for Windows | application/x-msworks | WPS, W40 | adWORDPROCESSOR |
| MS_Works_Win_DB_Fmt | 131 | 231 | Microsoft Works Database for Windows | application/x-msworks | | adDATABASE |
| MS_Works_Win_SS_Fmt | 132 | 228 | Microsoft Works Spreadsheet for Windows | application/x-msworks | S30, S40 | adSPREADSHEET |
| PC_Library_Fmt | 133 | 111 | DOS/Windows Object Library | application/x-archive | LIB, A | adLIBRARY |
| MacWrite_Fmt | 134 | 112 | MacWrite | application/macwriteii | | adWORDPROCESSOR |
| MacWrite_II_Fmt | 135 | 113 | MacWrite II | application/macwriteii | | adWORDPROCESSOR |
| Freehand_Fmt | 136 | 114 | Freehand MAC | image/x-freehand | | adVECTORGRAPHIC |
| Disk_Doubler_Fmt | 137 | 115 | Disk Doubler | | | adENCAPSULATION |
| HP_GL_Fmt | 138 | 116 | HP Graphics Language | vector/x-hpgl | HPGL | adVECTORGRAPHIC |
| FrameMaker_Fmt | 139 | 136 | FrameMaker | application/vnd.frameMaker | FM, FRM | adDESKTOPPUBLSH |
| FrameMaker_Book_Fmt | 140 | 136 | FrameMaker Book | application/vnd.frameMaker | BOOK | adDESKTOPPUBLSH |
| Maker_Markup_Language_Fmt | 141 | 174 | Maker Markup Language | application/vnd.mif | | adDESKTOPPUBLSH |
| Maker_Interchange_Fmt | 142 | 117 | Maker Interchange Format (MIF) | application/x-mif | MIF | adWORDPROCESSOR |
| JPEG_File_Interchange_Fmt | 143 | 118 | JPEG Interchange Format | image/jpeg | JPG, JPEG | adRASTERIMAGE |
| Reflex_Fmt | 144 | 119 | Borland Reflex database | | | adDATABASE |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|----------------------------|--------|----------|--|---------------------------|------------------|-----------------|
| Framework_Fmt | 145 | 276 | Framework office suite | | | adMIXED |
| Framework_II_Fmt | 146 | 120 | Framework II office suite | | FW3 | adMIXED |
| Paradox_Fmt | 147 | 121 | Borland Paradox database | | DB | adDATABASE |
| MS_Windows_Write_Fmt | 148 | 123 | Microsoft Windows Write | application/x-ms-write | WRI | adWORDPROCESSOR |
| Quattro_Pro_DOS_Fmt | 149 | 124 | Quattro Pro for DOS | application/x-quattropro | WQ1 | adSPREADSHEET |
| Quattro_Pro_Win_Fmt | 150 | 184 | Quattro Pro for Windows | application/x-quattro-win | WB1, WB2, WB3 | adSPREADSHEET |
| Persuasion_Fmt | 151 | 126 | Adobe Persuasion | | | adPRESENTATION |
| Windows_Icon_Fmt | 152 | 128 | Windows Icon Format | image/ico | ICO | adRASTERIMAGE |
| Windows_Cursor_Fmt | 153 | 133 | Windows Cursor | image/x-win-bitmap | CUR | adRASTERIMAGE |
| MS_Project_Activity_Fmt | 154 | 129 | Microsoft Project (up to version 3) activity file | | | adSCHEDULE |
| MS_Project_Resource_Fmt | 155 | 129 | Microsoft Project (up to version 3) resource file | | | adSCHEDULE |
| MS_Project_Calc_Fmt | 156 | 129 | Microsoft Project (up to version 3) calc file | | | adSCHEDULE |
| PKZIP_Fmt | 157 | 132 | ZIP Archive | application/zip | ZIP, ZIPX | adENCAPSULATION |
| Quark_Xpress_Fmt | 158 | 134 | Quark Xpress MAC | | | adDESKTOPPUBLSH |
| ARC_PAK_Archive_Fmt | 159 | 135 | PAK/ARC Archive | | ARC, PAK | adENCAPSULATION |
| MS_Publisher_Fmt | 160 | 137 | Microsoft Publisher (up to version 3) | application/x-mspublisher | PUB | adDESKTOPPUBLSH |
| PlanPerfect_Fmt | 161 | 138 | PlanPerfect | | | adSCHEDULE |
| WordPerfect_Auxiliary_Fmt | 162 | 139 | WordPerfect auxiliary file | | WPW | adMISC |
| MS_WAVE_Audio_Fmt | 163 | 141 | Microsoft Wave | audio/wav | WAV | adSOUND |
| MIDI_Audio_Fmt | 164 | 142 | MIDI audio | audio/mid | MID, MIDI | adSOUND |
| AutoCAD_DXF_Binary_Fmt | 165 | 143 | AutoCAD DXF | image/x-dxf | DXF | adVECTORGRAPHIC |
| AutoCAD_DXF_Text_Fmt | 166 | 143 | AutoCAD DXF | image/x-dxf | DXF | adVECTORGRAPHIC |
| dBase_Fmt | 167 | 144 | dBase | application/x-dbf | DBF, VCX | adDATABASE |
| OS_2_PM_Metafile_Fmt | 168 | 145 | OS/2 PM Metafile | | MET | adVECTORGRAPHIC |
| Lasergraphics_Language_Fmt | 169 | 146 | Lasergraphics Language | | | adVECTORGRAPHIC |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|---------------------------|--------|----------|---|--------------------------|-----------|-----------------|
| AutoShade_Rendering_Fmt | 170 | 147 | AutoShade Rendering | | | adVECTORGRAPHIC |
| GEM_VDI_Fmt | 171 | 148 | GEM VDI Metafile image | | GEM, GDI | adVECTORGRAPHIC |
| Windows_Help_Fmt | 172 | 149 | Windows Help File | application/winhelp | HLP | adMISC |
| Volkswriter_Fmt | 173 | 150 | Volkswriter word processor | | VW4 | adWORDPROCESSOR |
| Ability_WP_Fmt | 174 | 151 | Ability Word Processor | | | adWORDPROCESSOR |
| Ability_DB_Fmt | 175 | 151 | Ability Database | | | adDATABASE |
| Ability_SS_Fmt | 176 | 151 | Ability Spreadsheet | | | adSPREADSHEET |
| Ability_Comm_Fmt | 177 | 151 | Ability Presentation | | | adCOMMUNICATION |
| Ability_Image_Fmt | 178 | 151 | Ability Image | | | adRASTERIMAGE |
| XyWrite_Fmt | 179 | 152 | XYWrite / Nota Bene | | XY4 | adWORDPROCESSOR |
| CSV_Fmt | 180 | 153 | CSV (Comma Separated Values) | text/csv | CSV | adSPREADSHEET |
| IBM_Writing_Assistant_Fmt | 181 | 154 | IBM Writing Assistant | | IWA | adWORDPROCESSOR |
| WordStar_2000_Fmt | 182 | 155 | WordStar 2000 | | WS2 | adWORDPROCESSOR |
| HP_PCL_Fmt | 183 | 157 | HP Printer Control Language | application/pcl | PCL | adVECTORGRAPHIC |
| UNIX_Exe_PreSysV_VAX_Fmt | 184 | 158 | Unix Executable (PDP-11/pre-System V VAX) | application/octet-stream | | adEXECUTABLE |
| UNIX_Exe_Basic_16_Fmt | 185 | 158 | Unix Executable (Basic-16) | application/octet-stream | | adEXECUTABLE |
| UNIX_Exe_x86_Fmt | 186 | 158 | Unix Executable (x86) | application/octet-stream | | adEXECUTABLE |
| UNIX_Exe_iAPX_286_Fmt | 187 | 158 | Unix Executable (iAPX 286) | application/octet-stream | | adEXECUTABLE |
| UNIX_Exe_MC68k_Fmt | 188 | 158 | Unix Executable (MC680x0) | application/octet-stream | | adEXECUTABLE |
| UNIX_Exe_3B20_Fmt | 189 | 158 | Unix Executable (3B20) | application/octet-stream | | adEXECUTABLE |
| UNIX_Exe_WE32000_Fmt | 190 | 158 | Unix Executable (WE32000) | application/octet-stream | | adEXECUTABLE |
| UNIX_Exe_VAX_Fmt | 191 | 158 | Unix Executable (VAX) | application/octet-stream | | adEXECUTABLE |
| UNIX_Exe_Bell_5_Fmt | 192 | 158 | Unix Executable (Bell 5.0) | application/octet-stream | | adEXECUTABLE |
| UNIX_Obj_VAX_Demand_Fmt | 193 | 159 | Unix Object Module (VAX Demand) | | | adOBJECTMODULE |
| UNIX_Obj_MS8086_Fmt | 194 | 159 | Unix Object Module (old MS 8086) | | | adOBJECTMODULE |
| UNIX_Obj_Z8000_Fmt | 195 | 159 | Unix Object Module (Z8000) | | | adOBJECTMODULE |
| AU_Audio_Fmt | 196 | 161 | NeXT/Sun Audio Data | audio/basic | AU | adSOUND |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|-------------------------|--------|----------|--------------------------------|--------------------------|-----------|-----------------|
| NeWS_Font_Fmt | 197 | 162 | NeWS bitmap font | | | adFONT |
| cpio_Archive_CRCHdr_Fmt | 198 | 163 | cpio archive (CRC Header) | application/x-cpio | | adENCAPSULATION |
| cpio_Archive_CHRhdr_Fmt | 199 | 163 | cpio archive (CHR Header) | application/x-cpio | | adENCAPSULATION |
| PEX_Binary_Archive_Fmt | 200 | 164 | SUN PEX Binary Archive | | | adENCAPSULATION |
| Sun_vfont_Fmt | 201 | 165 | SUN vfont Definition | | | adFONT |
| Curses_Screen_Fmt | 202 | 166 | Curses Screen Image | | | adRASTERIMAGE |
| UUEncoded_Fmt | 203 | 167 | UU encoded | text/x-uencode | UUE | adENCAPSULATION |
| WriteNow_Fmt | 204 | 168 | WriteNow MAC | | | adWORDPROCESSOR |
| PC_Obj_Fmt | 205 | 169 | DOS/Windows Object Module | application/octet-stream | OBJ | adOBJECTMODULE |
| Windows_Group_Fmt | 206 | 170 | Windows Group | | | adMISC |
| TrueType_Font_Fmt | 207 | 171 | TrueType Font | application/x-font-ttf | TTF | adFONT |
| Windows_PIF_Fmt | 208 | 172 | Program Information File (PIF) | application/octet-stream | PIF | adMISC |
| MS_COM_Executable_Fmt | 209 | 173 | PC (.COM) | application/octet-stream | COM | adEXECUTABLE |
| Stuftit_Fmt | 210 | 175 | Stuftit (MAC) | application/x-stuftit | HQX | adENCAPSULATION |
| PeachCalc_Fmt | 211 | 176 | PeachCalc | | CAL | adSPREADSHEET |
| Wang_GDL_Fmt | 212 | 177 | WANG Office GDL Header | | | adENCAPSULATION |
| Q_A_DOS_Fmt | 213 | 179 | Q & A for DOS | | | adWORDPROCESSOR |
| Q_A_Win_Fmt | 214 | 180 | Q & A for Windows | | JW | adWORDPROCESSOR |
| WPS_PLUS_Fmt | 215 | 181 | WPS-PLUS | application/vnd.ms-wpl | WPL | adWORDPROCESSOR |
| DCX_Fmt | 216 | 182 | DCX FAX Format(PCX images) | image/dcx | DCX | adFAXFORMAT |
| OLE_Fmt | 217 | 183 | OLE Compound Document | | OLE | adENCAPSULATION |
| EBCDIC_Fmt | 218 | 186 | EBCDIC Text | | | adWORDPROCESSOR |
| DCS_Fmt | 219 | 187 | DCS | | | adWORDPROCESSOR |
| UNIX_SHAR_Fmt | 220 | 190 | SHAR shell archive format | application/x-shar | SHAR | adENCAPSULATION |
| Lotus_Notes_BitMap_Fmt | 221 | 191 | Lotus Notes Bitmap | | | adRASTERIMAGE |
| Lotus_Notes_CDF_Fmt | 222 | 193 | Lotus Notes CDF | application/cdf | CDF | adWORDPROCESSOR |
| Compress_Fmt | 223 | 192 | Unix Compress | application/x-compress | Z | adENCAPSULATION |
| GZ_Compress_Fmt | 224 | 198 | GZ Compress | application/gzip | GZ | adENCAPSULATION |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|-----------------------------|--------|----------|---|--------------------------|--------------|-----------------|
| TAR_Fmt | 225 | 194 | TAR archive | application/tar | TAR | adENCAPSULATION |
| ODIF_FOD26_Fmt | 226 | 196 | Open Document Architecture (ODA / ODIF) FOD26 | application/oda | F26 | adWORDPROCESSOR |
| ODIF_FOD36_Fmt | 227 | 196 | Open Document Architecture (ODA / ODIF) FOD36 | application/oda | F36 | adWORDPROCESSOR |
| ALIS_Fmt | 228 | 197 | ALIS | | | adWORDPROCESSOR |
| Envoy_Fmt | 229 | 199 | WordPerfect Envoy | application/envoy | EVY | adWORDPROCESSOR |
| PDF_Fmt | 230 | 200 | Portable Document Format | application/pdf | PDF | adWORDPROCESSOR |
| BinHex_Fmt | 231 | 206 | BinHex | application/mac-binhex40 | HQX | adENCAPSULATION |
| SMTP_Fmt | 232 | 207 | SMTP | message/rfc822 | SMTP | adENCAPSULATION |
| MIME_Fmt | 233 | 208 | MIME (EML, MBX email) ¹ | message/rfc822 | EML, MBX | adENCAPSULATION |
| USENET_Fmt | 234 | 264 | USENET | message/news | | adWORDPROCESSOR |
| SGML_Fmt | 235 | 209 | SGML | text/sgml | SGML | adWORDPROCESSOR |
| HTML_Fmt | 236 | 210 | HTML | text/html | HTM, HTML | adWORDPROCESSOR |
| ACT_Fmt | 237 | 211 | ACT! CRM software | | ACT | adWORDPROCESSOR |
| PNG_Fmt | 238 | 213 | Portable Network Graphics (PNG) | image/png | PNG | adRASTERIMAGE |
| MS_Video_Fmt | 239 | 214 | Video for Windows (AVI) | video/avi | AVI | adMOVIE |
| Windows_Animated_Cursor_Fmt | 240 | 215 | Windows Animated Cursor | | ANI | adRASTERIMAGE |
| Windows_CPP_Obj_Storage_Fmt | 241 | 216 | Windows C++ Object Storage | | | adMIXED |
| Windows_Palette_Fmt | 242 | 217 | Windows Palette | | PAL | adRASTERIMAGE |
| RIFF_DIB_Fmt | 243 | 218 | RIFF Device Independent Bitmap | | | adRASTERIMAGE |
| RIFF_MIDI_Fmt | 244 | 219 | RIFF MIDI | audio/midi | RMI | adSOUND |
| RIFF_Multimedia_Movie_Fmt | 245 | 220 | RIFF Multimedia Movie | | | adMOVIE |
| MPEG_Fmt | 246 | 221 | MPEG Movie | video/mpeg | | adMOVIE |
| QuickTime_Fmt | 247 | 222 | QuickTime Movie, MPEG-4 audio | video/quicktime | MOV, QT, MP4 | adMOVIE |
| AIFF_Fmt | 248 | 223 | Audio Interchange File Format (AIFF) | audio/aiff | AIF, AIFF | adSOUND |
| Amiga_MOD_Fmt | 249 | 224 | Amiga MOD | | MOD | adSOUND |
| Amiga_IFF_8SVX_Fmt | 250 | 225 | Amiga IFF (8SVX) Sound | audio/x-8svx | IFF | adSOUND |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|------------------------------|--------|----------|----------------------------------|---------------------------------|-----------------|-----------------|
| Creative_Voice_Audio_Fmt | 251 | 226 | Creative Voice (VOC) | | VOC | adSOUND |
| AutoDesk_Animator_FLI_Fmt | 252 | 229 | AutoDesk Animator FLIC | video/x-flc | FLI | adANIMATION |
| AutoDesk_AnimatorPro_FLC_Fmt | 253 | 230 | AutoDesk Animator Pro FLIC | video/x-flc | FLC | adANIMATION |
| Compactor_Archive_Fmt | 254 | 233 | Compactor / Compact Pro | application/mac-compactpro | | adENCAPSULATION |
| VRML_Fmt | 255 | 234 | VRML | model/vrml | WRL | adVECTORGRAPHIC |
| QuickDraw_3D_Metafile_Fmt | 256 | 235 | QuickDraw 3D Metafile | | | adVECTORGRAPHIC |
| PGP_Secret_Keyring_Fmt | 257 | 236 | PGP Secret Keyring | application/pgp | | adENCAPSULATION |
| PGP_Public_Keyring_Fmt | 258 | 237 | PGP Public Keyring | application/pgp | | adENCAPSULATION |
| PGP_Encrypted_Data_Fmt | 259 | 238 | PGP Encrypted Data | application/pgp | | adENCAPSULATION |
| PGP_Signed_Data_Fmt | 260 | 239 | PGP Signed Data | application/pgp | | adENCAPSULATION |
| PGP_SignedEncrypted_Data_Fmt | 261 | 240 | PGP Signed and Encrypted Data | application/pgp | | adENCAPSULATION |
| PGP_Sign_Certificate_Fmt | 262 | 241 | PGP Signature Certificate | application/pgp-signature | SIG | adENCAPSULATION |
| PGP_Compressed_Data_Fmt | 263 | 246 | PGP Compressed Data | application/pgp | | adENCAPSULATION |
| PGP_ASCII_Public_Keyring_Fmt | 264 | 242 | ASCII-armored PGP Public Keyring | application/pgp | PGP | adENCAPSULATION |
| PGP_ASCII_Encoded_Fmt | 265 | 243 | ASCII-armored PGP encoded | application/pgp | | adENCAPSULATION |
| PGP_ASCII_Signed_Fmt | 266 | 244 | ASCII-armored PGP signed | application/pgp | | adENCAPSULATION |
| OLE_DIB_Fmt | 267 | 245 | OLE DIB object | | | adRASTERIMAGE |
| SGI_Image_Fmt | 268 | 247 | SGI Image | image/sgi | RGB | adRASTERIMAGE |
| Lotus_ScreenCam_Fmt | 269 | 248 | Lotus ScreenCam | application/vnd.lotus-screencam | SCM | adANIMATION |
| MPEG_Audio_Fmt | 270 | 249 | MPEG Audio | audio/mpeg | MPEGA, MPG, MP3 | adSOUND |
| FTP_Software_Session_Fmt | 271 | 250 | FTP Session Data | | STE | adCOMMUNICATION |
| Netscape_Bookmark_File_Fmt | 272 | 210 | Netscape Bookmark File | text/html | | adWORDPROCESSOR |
| Corel_Draw_CMx_Fmt | 273 | 252 | Corel CMX | application/cmx | CMX | adVECTORGRAPHIC |
| AutoDesk_DWG_Fmt | 274 | 253 | AutoDesk Drawing (DWG) | image/x-dwg | DWG | adVECTORGRAPHIC |
| AutoDesk_WHIP_Fmt | 275 | 254 | AutoDesk WHIP | | WHP | adVECTORGRAPHIC |
| Macromedia_Director_Fmt | 276 | 255 | Macromedia Director | application/x-director | DCR | adANIMATION |
| Real_Audio_Fmt | 277 | 256 | Real Audio | audio/x-pn-realaudio | RM, RA | adSOUND |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|-------------------------|--------|----------|-------------------------------------|-------------------------------|--------------------|-----------------|
| MSDOS_Device_Driver_Fmt | 278 | 257 | MSDOS Device Driver | application/octet-stream | SYS | adEXECUTABLE |
| Micrografx_Designer_Fmt | 279 | 258 | Micrografx Designer | | DSF | adVECTORGRAPHIC |
| SVF_Fmt | 280 | 259 | Simple Vector Format (SVF) | image/x-svf | SVF | adVECTORGRAPHIC |
| Applix_Words_Fmt | 281 | 261 | Applix Words | application/x-applix-word | AW | adWORDPROCESSOR |
| Applix_Graphics_Fmt | 282 | 262 | Applix Graphics | | AG | adPRESENTATION |
| MS_Access_Fmt | 283 | 263 | Microsoft Access (versions 1 and 2) | application/x-msaccess | MDB | adDATABASE |
| MS_Access_95_Fmt | 284 | 263 | Microsoft Access 95 | application/msaccess | MDB | adDATABASE |
| MS_Access_97_Fmt | 285 | 263 | Microsoft Access 97 | application/msaccess | MDB | adDATABASE |
| MacBinary_Fmt | 286 | 265 | MacBinary | application/x-macbinary | BIN | adENCAPSULATION |
| Apple_Single_Fmt | 287 | 266 | Apple Single | | | adENCAPSULATION |
| Apple_Double_Fmt | 288 | 267 | Apple Double | multipart/appledouble | AD | adENCAPSULATION |
| Enhanced_Metafile_Fmt | 289 | 270 | Enhanced Metafile | image/x-emf | EMF | adVECTORGRAPHIC |
| MS_Office_Drawing_Fmt | 290 | 271 | Microsoft Office Drawing | | | adVECTORGRAPHIC |
| XML_Fmt | 291 | 285 | XML | text/xml | XML | adWORDPROCESSOR |
| DeVice_Independent_Fmt | 292 | 274 | DeVice Independent file (DVI) | application/x-dvi | DVI | adVECTORGRAPHIC |
| Unicode_Fmt | 293 | 275 | Unicode text file | text/plain | UNI | adWORDPROCESSOR |
| Lotus_123_Worksheet_Fmt | 294 | 81 | Lotus 1-2-3 | application/x-lotus-123 | WKS, WK1, WK3, WK4 | adSPREADSHEET |
| Lotus_123_Format_Fmt | 295 | 81 | Lotus 1-2-3 Formatting | application/x-123 | FM3 | adSPREADSHEET |
| Lotus_123_97_Fmt | 296 | 81 | Lotus 1-2-3 97 | application/x-lotus-123 | 123 | adSPREADSHEET |
| Lotus_Word_Pro_96_Fmt | 297 | 268 | Lotus Word Pro 96 | application/vnd.lotus-wordpro | LWP, MWP | adWORDPROCESSOR |
| Lotus_Word_Pro_97_Fmt | 298 | 268 | Lotus Word Pro 97 | application/vnd.lotus-wordpro | LWP, MWP | adWORDPROCESSOR |
| Freelance_DOS_Fmt | 299 | 140 | Lotus Freelance for DOS | application/x-freelance | PRZ | adPRESENTATION |
| Freelance_Win_Fmt | 300 | 140 | Lotus Freelance for Windows | application/x-freelance | PRE | adPRESENTATION |
| Freelance_OS2_Fmt | 301 | 140 | Lotus Freelance for OS/2 | application/x-freelance | PRS | adPRESENTATION |
| Freelance_96_Fmt | 302 | 140 | Lotus Freelance 96 | application/x-freelance | PRZ | adPRESENTATION |
| Freelance_97_Fmt | 303 | 140 | Lotus Freelance 97 | application/x-freelance | PRZ | adPRESENTATION |
| MS_Word_95_Fmt | 304 | 189 | Microsoft Word 95 | application/msword | DOC | adWORDPROCESSOR |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|------------------------------|--------|----------|---|----------------------------------|---------------|-----------------|
| MS_Word_97_Fmt | 305 | 269 | Microsoft Word 97 | application/msword | DOC, WPS, WBK | adWORDPROCESSOR |
| Excel_Fmt | 306 | 90 | Microsoft Excel (up to version 5) | application/x-ms-excel | XLS | adSPREADSHEET |
| Excel_Chart_Fmt | 307 | 90 | Microsoft Excel (up to version 5) chart | application/x-ms-excel | XLC | adSPREADSHEET |
| Excel_Macro_Fmt | 308 | 90 | Microsoft Excel (up to version 5) macro | application/vnd.ms-excel | XLM | adSPREADSHEET |
| Excel_95_Fmt | 309 | 188 | Microsoft Excel 95 | application/x-ms-excel | XLS | adSPREADSHEET |
| Excel_97_Fmt | 310 | 188 | Microsoft Excel 97 | application/x-ms-excel | XLS | adSPREADSHEET |
| Corel_Presentations_Fmt | 311 | 127 | Corel Presentations | application/x-corelpresentations | XFD, XFDL | adPRESENTATION |
| Harvard_Graphics_Fmt | 312 | 131 | Harvard Graphics | | PR4 | adPRESENTATION |
| Harvard_Graphics_Chart_Fmt | 313 | 131 | Harvard Graphics Chart | | CH3, CHT | adVECTORGRAPHIC |
| Harvard_Graphics_Symbol_Fmt | 314 | 131 | Harvard Graphics Symbol File | | SY3 | adVECTORGRAPHIC |
| Harvard_Graphics_Cfg_Fmt | 315 | 131 | Harvard Graphics Configuration File | | | adVECTORGRAPHIC |
| Harvard_Graphics_Palette_Fmt | 316 | 131 | Harvard Graphics Palette | | | adVECTORGRAPHIC |
| Lotus_123_R9_Fmt | 317 | 81 | Lotus 1-2-3 Release 9 | application/x-lotus-123 | 123 | adSPREADSHEET |
| Applix_Spreadsheets_Fmt | 318 | 278 | Applix Spreadsheets | application/x-applix-spreadsheet | AS | adSPREADSHEET |
| MS_Pocket_Word_Fmt | 319 | 45 | Microsoft Pocket Word | | PWD | adWORDPROCESSOR |
| MS_DIB_Fmt | 320 | 279 | Microsoft Device Independent Bitmap | image/bmp | DIB | adRASTERIMAGE |
| MS_Word_2000_Fmt | 321 | 269 | Microsoft Word 2000 | application/msword | DOC | adWORDPROCESSOR |
| Excel_2000_Fmt | 322 | 188 | Microsoft Excel 2000 | application/x-ms-excel | XLS | adSPREADSHEET |
| PowerPoint_2000_Fmt | 323 | 272 | Microsoft PowerPoint 2000 | application/x-ms-powerpoint | PPT | adPRESENTATION |
| MS_Access_2000_Fmt | 324 | 263 | Microsoft Access 2000 | application/x-msaccess | MDB | adDATABASE |
| MS_Project_4_Fmt | 325 | 281 | Microsoft Project 4 | | MPP | adSCHEDULE |
| MS_Project_41_Fmt | 326 | 281 | Microsoft Project 4.1 | | MPP | adSCHEDULE |
| MS_Project_98_Fmt | 327 | 281 | Microsoft Project 98 | application/vnd.ms-project | MPP | adSCHEDULE |
| Folio_Flat_Fmt | 328 | 282 | Folio Flat File | | FFF | adWORDPROCESSOR |
| HWP_Fmt | 329 | 283 | HWP (Arae-Ah Hangul) | application/x-hwp | HWP | adWORDPROCESSOR |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|------------------------|--------|----------|--|--------------------------------|---------------|-----------------|
| ICHITARO_Fmt | 330 | 284 | ICHITARO (v4-10) | | JTD | adWORDPROCESSOR |
| IS_XML_Fmt | 331 | 273 | Extended or Custom XML | text/xml | XML | adWORDPROCESSOR |
| Oasys_Fmt | 332 | 286 | Oasys | application/vnd.fujitsu.oasys | OAS, OA2, OA3 | adWORDPROCESSOR |
| PBM_ASC_Fmt | 333 | 287 | Portable Bitmap Utilities ASCII format (PBM) | image/pbm | PBM | adRASTERIMAGE |
| PBM_BIN_Fmt | 334 | 287 | Portable Bitmap Utilities BINARY format (PBM) | image/pbm | PBM | adRASTERIMAGE |
| PGM_ASC_Fmt | 335 | 288 | Portable Greymap Utilities ASCII format (PGM) | image/x-pgm | PGM | adRASTERIMAGE |
| PGM_BIN_Fmt | 336 | 288 | Portable Greymap Utilities BINARY format (PGM) | image/x-pgm | PGM | adRASTERIMAGE |
| PPM_ASC_Fmt | 337 | 289 | Portable Pixmap Utilities ASCII format (PPM) | image/x-portable-pixmap | PPM | adRASTERIMAGE |
| PPM_BIN_Fmt | 338 | 289 | Portable Pixmap Utilities BINARY format (PPM) | image/x-portable-pixmap | PPM | adRASTERIMAGE |
| XBM_Fmt | 339 | 290 | X Bitmap format (XBM) | image/x-xbitmap | XBM | adRASTERIMAGE |
| XPM_Fmt | 340 | 291 | X Pixmap format (XPM) | image/xpm | XPM | adRASTERIMAGE |
| FPX_Fmt | 341 | 292 | Kodak FlashPix FPX Image format | image/fpx | FPX | adRASTERIMAGE |
| PCD_Fmt | 342 | 293 | PCD Image format | image/pcd | PCD | adRASTERIMAGE |
| MS_Visio_Fmt | 343 | 294 | Microsoft Visio (up to version 11) | image/x-vsd | VSD | adPRESENTATION |
| MS_Project_2000_Fmt | 344 | 281 | Microsoft Project 2000 | application/vnd.ms-project | MPP | adSCHEDULE |
| MS_Outlook_Fmt | 345 | 295 | Microsoft Outlook message | application/vnd.ms-outlook | MSG, OFT | adENCAPSULATION |
| ELF_Relocatable_Fmt | 346 | 159 | ELF Relocatable | application/octet-stream | O | adOBJECTMODULE |
| ELF_Executable_Fmt | 347 | 158 | ELF Executable | application/octet-stream | | adEXECUTABLE |
| ELF_Dynamic_Lib_Fmt | 348 | 160 | ELF Dynamic Library | application/octet-stream | SO | adLIBRARY |
| MS_Word_XML_Fmt | 349 | 285 | Microsoft Word 2003 XML | text/xml | XML | adWORDPROCESSOR |
| MS_Excel_XML_Fmt | 350 | 285 | Microsoft Excel 2003 XML | text/xml | XML | adWORDPROCESSOR |
| MS_Visio_XML_Fmt | 351 | 285 | Microsoft Visio 2003 XML | text/xml | VDX | adWORDPROCESSOR |
| SO_Text_XML_Fmt | 352 | 314 | OpenDocument format (OpenOffice 1/StarOffice 6,7) Text XML | application/vnd.sun.xml.writer | SXW | adWORDPROCESSOR |
| SO_Spreadsheet_XML_Fmt | 353 | 315 | OpenDocument format | application/vnd.sun.xml.calc | SXC, STC | adSPREADSHEET |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|-------------------------|--------|----------|--|---|---------------------------|-----------------|
| | | | (OpenOffice 1/StarOffice 6,7) Spreadsheet XML | | | |
| SO_Presentation_XML_Fmt | 354 | 316 | OpenDocument format (OpenOffice 1/StarOffice 6,7) Presentation XML | application/vnd.sun.xml.impress | SXD, SXI | adPRESENTATION |
| XHTML_Fmt | 355 | 296 | XHTML | text/xhtml | XML, ASP | adWORDPROCESSOR |
| MS_OutlookPST_Fmt | 356 | 297 | Microsoft Outlook Personal Folders File (.pst) | application/vnd.ms-outlook-pst | PST | adENCAPSULATION |
| RAR_Fmt | 357 | 298 | RAR archive format | application/x-rar-compressed | RAR | adENCAPSULATION |
| Lotus_Notes_NSF_Fmt | 358 | 299 | IBM Lotus Notes Database NSF/NTF | application/x-lotus-notes | NSF | adENCAPSULATION |
| Macromedia_Flash_Fmt | 359 | 300 | Macromedia Flash (.swf) | application/x-shockwave-flash | SWF | adWORDPROCESSOR |
| MS_Word_2007_Fmt | 360 | 301 | Microsoft Word 2007 XML - Docx | application/x-ms-word07 | DOCX, DOTX | adWORDPROCESSOR |
| MS_Excel_2007_Fmt | 361 | 302 | Microsoft Excel 2007 XML | application/x-ms-excel07 | XLSX, XLTX | adSPREADSHEET |
| MS_PPT_2007_Fmt | 362 | 303 | Microsoft PowerPoint 2007 XML | application/x-ms-powerpoint07 | PPTX, POTX, PPSX | adPRESENTATION |
| OpenPGP_Fmt | 363 | 304 | OpenPGP Message Format (with new packet format) | application/pgp-encrypted | PGP | adENCAPSULATION |
| Intergraph_V7_DGN_Fmt | 364 | 305 | Intergraph Standard File Format (ISFF) V7 DGN (non-OLE) | | DGN | adVECTORGRAPHIC |
| MicroStation_V8_DGN_Fmt | 365 | 306 | MicroStation V8 DGN (OLE) | | DGN | adVECTORGRAPHIC |
| MS_Word_Macro_2007_Fmt | 366 | 307 | Microsoft Word Macro 2007 XML | application/x-ms-word07m | DOCM, DOTM | adWORDPROCESSOR |
| MS_Excel_Macro_2007_Fmt | 367 | 308 | Microsoft Excel Macro 2007 XML | application/x-ms-excel07m | XLSM, XLTM, XLAM | adSPREADSHEET |
| MS_PPT_Macro_2007_Fmt | 368 | 309 | Microsoft PPT Macro 2007 XML | application/x-ms-powerpoint07m | PPTM, POTM, PPSM, PPAM | adPRESENTATION |
| LZH_Fmt | 369 | 310 | LZH Archive | application/x-lzh-compressed | LZH, LHA | adENCAPSULATION |
| Office_2007_Fmt | 370 | 311 | Office 2007 document | | XLSB | adMISC |
| MS_XPS_Fmt | 371 | 312 | Microsoft XML Paper Specification (XPS) | application/vnd.ms-xpsdocument | XPS | adWORDPROCESSOR |
| Lotus_Domino_DXL_Fmt | 372 | 313 | IBM Domino Data in XML format (.dxl) | text/xml | DXL | adENCAPSULATION |
| ODF_Text_Fmt | 373 | 314 | ODF Text | application/vnd.oasis.opendocument.text | ODT | adWORDPROCESSOR |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|---------------------------|--------|----------|---|---|-----------------------------|-----------------|
| ODF_Spreadsheet_Fmt | 374 | 315 | ODF Spreadsheet | application/vnd.oasis.opendocument.spreadsheet | ODS | adSPREADSHEET |
| ODF_Presentation_Fmt | 375 | 316 | ODF Presentation | application/vnd.oasis.opendocument.presentation | ODP | adPRESENTATION |
| Legato_Extender_ONM_Fmt | 376 | 317 | Legato Extender Native Message ONM | application/x-lotus-notes | ONM | adENCAPSULATION |
| bin_Unknown_Fmt | 377 | 318 | Bin unknown format (.xxx) | | | adWORDPROCESSOR |
| TNEF_Fmt | 378 | 319 | Transport Neutral Encapsulation Format (TNEF) | application/vnd.ms-tnef | | adENCAPSULATION |
| CADAM_Drawing_Fmt | 379 | 320 | CADAM Drawing | | CDD | adVECTORGRAPHIC |
| CADAM_Drawing_Overlay_Fmt | 380 | 321 | CADAM Drawing Overlay | | CDO | adVECTORGRAPHIC |
| NURSTOR_Drawing_Fmt | 381 | 322 | NURSTOR Drawing | | NUR | adVECTORGRAPHIC |
| HP_GLP_Fmt | 382 | 323 | HP Graphics Language (Plotter) | vector/x-hpgl2 | HPG | adVECTORGRAPHIC |
| ASF_Fmt | 383 | 324 | Advanced Systems Format (ASF) | application/x-ms-asf | ASF | adMISC |
| WMA_Fmt | 384 | 325 | Windows Media Audio Format (WMA) | audio/x-ms-wma | WMA | adSOUND |
| WMV_Fmt | 385 | 326 | Windows Media Video Format (WMV) | video/x-ms-wmv | WMV | adMOVIE |
| EMX_Fmt | 386 | 327 | Legato EMailXtender Archives Format (EMX) | | EMX | adENCAPSULATION |
| Z7Z_Fmt | 387 | 328 | 7 Zip Format (7z) | application/7z | 7Z | adENCAPSULATION |
| MS_Excel_Binary_2007_Fmt | 388 | 329 | Microsoft Excel Binary 2007 | application/vnd.ms-excel.sheet.binary.macroenabled.12 | XLSB | adSPREADSHEET |
| CAB_Fmt | 389 | 330 | Microsoft Cabinet File (CAB) | application/vnd.ms-cab-compressed | CAB | adENCAPSULATION |
| CATIA_Fmt | 390 | 331 | CATIA Formats (CAT*) | | CATPART, CATPRODUCT 2 | adVECTORGRAPHIC |
| YIM_Fmt | 391 | 332 | Yahoo Instant Messenger History | | DAT | adWORDPROCESSOR |
| ODF_Drawing_Fmt | 392 | 316 | ODF Drawing/Graphics | application/vnd.oasis.opendocument.graphics | ODG | adVECTORGRAPHIC |
| Founder_CEB_Fmt | 393 | 333 | Founder Chinese E-paper Basic (ceb) | application/ceb | CEB | adWORDPROCESSOR |
| QPW_Fmt | 394 | 334 | Corel Quattro Pro 9+ for Windows | application/quattro-pro | QPW | adSPREADSHEET |
| MHT_Fmt | 395 | 335 | MHTML format (MHT) ¹ | multipart/related | MHT, MHTML | adWORDPROCESSOR |
| MDI_Fmt | 396 | 336 | Microsoft Document Imaging Format | image/vnd.ms-modi | MDI | adRASTERIMAGE |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|------------------------|--------|----------|---------------------------------------|-------------------------------------|---|-----------------|
| GRV_Fmt | 397 | 337 | Microsoft Office Groove Format | application/vnd.groove-injector | GRV | adWORDPROCESSOR |
| IWWP_Fmt | 398 | 338 | Apple iWork Pages format | application/vnd.apple.pages | PAGES | adWORDPROCESSOR |
| IWSS_Fmt | 399 | 339 | Apple iWork Numbers format | application/vnd.apple.numbers | NUMBERS | adSPREADSHEET |
| IWPG_Fmt | 400 | 340 | Apple iWork Keynote format | application/vnd.apple.keynote | KEY | adPRESENTATION |
| BKF_Fmt | 401 | 341 | Windows Backup File | | BKF | adENCAPSULATION |
| MS_Access_2007_Fmt | 402 | 342 | Microsoft Access 2007 | application/msaccess | ACCDB | adDATABASE |
| ENT_Fmt | 403 | 343 | Microsoft Entourage Database Format | | | adENCAPSULATION |
| DMG_Fmt | 404 | 344 | Mac Disk Copy Disk Image File | application/x-apple-diskimage | DMG | adENCAPSULATION |
| CWK_Fmt | 405 | 345 | AppleWorks File | application/appleworks | CWK | adWORDPROCESSOR |
| OO3_Fmt | 406 | 346 | Omni Outliner V3 File | | OO3 | adWORDPROCESSOR |
| OPML_Fmt | 407 | 347 | Omni Outliner OPML File | | OPML | adWORDPROCESSOR |
| Omni_Graffle_XML_Fmt | 408 | 348 | Omni Graffle XML File | | GRAFFLE | adVECTORGRAPHIC |
| PSD_Fmt | 409 | 349 | Photoshop Document | image/vnd.adobe.photoshop | PSD, PSB | adRASTERIMAGE |
| Apple_Binary_PList_Fmt | 410 | 350 | Apple Binary Property List format | | PLIST | adMISC |
| Apple_iChat_Fmt | 411 | 351 | Apple iChat format | | ICHAT | adWORDPROCESSOR |
| OOOUTLINE_Fmt | 412 | 352 | OOutliner File | | OOOUTLINE | adWORDPROCESSOR |
| BZIP2_Fmt | 413 | 353 | Bzip 2 Compressed File | application/x-bzip2 | BZ2 | adENCAPSULATION |
| ISO_Fmt | 414 | 354 | ISO-9660 CD Disc Image Format | application/x-iso9660-image | ISO | adENCAPSULATION |
| DocuWorks_Fmt | 415 | 355 | DocuWorks Format | application/vnd.fujixerox.docuworks | XDW | adWORDPROCESSOR |
| RealMedia_Fmt | 416 | 356 | RealMedia Streaming Media | application/vnd.rn-realmedia | RM, RA | adMOVIE |
| AC3Audio_Fmt | 417 | 357 | AC3 Audio File Format | audio/ac3 | AC3 | adSOUND |
| NEF_Fmt | 418 | 358 | Nero Encrypted File | | NEF | adENCAPSULATION |
| SolidWorks_Fmt | 419 | 359 | SolidWorks Format Files | | SLDASM, SLDPRT, SLDDRW, SLDDRT | adVECTORGRAPHIC |
| XFDL_Fmt | 420 | 366 | Extensible Forms Description Language | application/x-xfdl | XFDL, XFD | adPRESENTATION |
| Apple_XML_PList_Fmt | 421 | 367 | Apple XML Property List format | | PLIST | adMISC |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|------------------------|--------|----------|--|--------------------------------|---|-----------------|
| OneNote_Fmt | 422 | 368 | OneNote Note Format | application/onenote | ONE | adWORDPROCESSOR |
| IFilter_Fmt | 423 | 369 | iFilter | | | adWORDPROCESSOR |
| Dicom_Fmt | 424 | 370 | Digital Imaging and Communications in Medicine (Dicom) | application/dicom | DCM | adRASTERIMAGE |
| EnCase_Fmt | 425 | 371 | Expert Witness Compression Format (EnCase) | | E01, L01, Lx01 | adENCAPSULATION |
| Scrap_Fmt | 426 | 372 | Shell Scrap Object File | | SHS | adENCAPSULATION |
| MS_Project_2007_Fmt | 427 | 373 | Microsoft Project 2007 | application/vnd.ms-project | MPP | adSCHEDULE |
| MS_Publisher_98_Fmt | 428 | 374 | Microsoft Publisher from version 98 | application/x-mspublisher | PUB | adDESKTOPPUBLSH |
| Skype_Fmt | 429 | 375 | Skype Log File | | DBB | adWORDPROCESSOR |
| HL7_Fmt | 430 | 377 | Health level7 message | | HL7 | adWORDPROCESSOR |
| MS_OutlookOST_Fmt | 431 | 378 | Microsoft Outlook Offline Folders File (OST) | application/vnd.ms-outlook-pst | OST | adENCAPSULATION |
| Epub_Fmt | 432 | 379 | Electronic Publication | application/epub+zip | EPUB | adWORDPROCESSOR |
| MS_OEDBX_Fmt | 433 | 380 | Microsoft Outlook Express DBX Message Database | | DBX | adENCAPSULATION |
| BB_Activ_Fmt | 434 | 381 | BlackBerry Activation File | | DAT | adWORDPROCESSOR |
| DiskImage_Fmt | 435 | 382 | Disk Image | | DMG | adENCAPSULATION |
| Milestone_Fmt | 436 | 383 | Milestone Document | | MLS, ML3, ML4, ML5, ML6, ML7, ML8, ML9, MLA | adRASTERIMAGE |
| E_Transcript_Fmt | 437 | 384 | RealLegal E-Transcript File | | PTX | adWORDPROCESSOR |
| PostScript_Font_Fmt | 438 | 385 | PostScript Type 1 Font | application/x-font | PFB | adFONT |
| Ghost_DiskImage_Fmt | 439 | 386 | Ghost Disk Image File | | GHO, GHS | adENCAPSULATION |
| JPEG_2000_JP2_File_Fmt | 440 | 387 | JPEG-2000 JP2 File Format Syntax (ISO/IEC 15444-1) | image/jp2 | JP2, JPF, J2K, JPWL, JPX, PGX | adRASTERIMAGE |
| Unicode_HTML_Fmt | 441 | 388 | Unicode HTML | text/html | HTM, HTML | adWORDPROCESSOR |
| CHM_Fmt | 442 | 389 | Microsoft Compiled HTML Help | application/x-chm | CHM | adENCAPSULATION |
| EMCMF_Fmt | 443 | 390 | Documentum EMCMF format | | EMCMF | adENCAPSULATION |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|-------------------------|--------|----------|---|---------------------------------------|-------------|-----------------|
| MS_Access_2007_Tmpl_Fmt | 444 | 391 | Microsoft Access 2007 Template | | ACCDT | adDATABASE |
| Jungum_Fmt | 445 | 392 | Samsung Electronics Jungum Global document | | GUL | adWORDPROCESSOR |
| JBIG2_Fmt | 446 | 393 | JBIG2 File Format | image/jbig2 | JB2, JBIG2 | adRASTERIMAGE |
| EFax_Fmt | 447 | 394 | eFax file | | EFX | adRASTERIMAGE |
| AD1_Fmt | 448 | 395 | AD1 Evidence file | | AD1 | adENCAPSULATION |
| SketchUp_Fmt | 449 | 396 | Google SketchUp | | SKP | adVECTORGRAPHIC |
| GWFS_Email_Fmt | 450 | 397 | Group Wise File Surf email | | GWFS | adENCAPSULATION |
| JNT_Fmt | 451 | 398 | Windows Journal format | | JNT | adWORDPROCESSOR |
| Yahoo_yChat_Fmt | 452 | 399 | Yahoo! Messenger chat log | | YCHAT | adWORDPROCESSOR |
| PaperPort_MAX_File_Fmt | 453 | 400 | PaperPort MAX image file | image/max | MAX | adRASTERIMAGE |
| ARJ_Fmt | 454 | 402 | ARJ (Archive by Robert Jung) file format | application/arj | ARJ | adENCAPSULATION |
| RPMSG_Fmt | 455 | 403 | Microsoft Outlook Restricted Permission Message | application/x-microsoft-rpmsg-message | RPMSG | adENCAPSULATION |
| MAT_Fmt | 456 | 404 | MATLAB file format | application/x-matlab-data | MAT, FIG | adWORDPROCESSOR |
| SGY_Fmt | 457 | 405 | SEG-Y Seismic Data format | | SGY, SEGY | adWORDPROCESSOR |
| CDXA_MPEG_PS_Fmt | 458 | 406 | MPEG-PS container with CDXA stream | video/mpeg | MPG | adMOVIE |
| EVT_Fmt | 459 | 407 | Microsoft Windows NT Event Log | | EVT | adMISC |
| EVTX_Fmt | 460 | 408 | Microsoft Windows Vista Event Log | | EVTX | adMISC |
| MS_OutlookOLM_Fmt | 461 | 409 | Microsoft Outlook for Macintosh format | | OLM | adENCAPSULATION |
| WARC_Fmt | 462 | 410 | Web ARChive | application/warc | WARC | adENCAPSULATION |
| JAVAClass_Fmt | 463 | 411 | Java Class format | application/x-java-class | CLASS | adWORDPROCESSOR |
| VCF_Fmt | 464 | 412 | Microsoft Outlook vCard file format | text/vcard | VCF | adWORDPROCESSOR |
| EDB_Fmt | 465 | 413 | Microsoft Exchange Server Database file format | | EDB | adENCAPSULATION |
| ICS_Fmt | 466 | 414 | Microsoft Outlook iCalendar file format | text/calendar | ICS, VCS | adENCAPSULATION |
| MS_Visio_2013_Fmt | 467 | 415 | Microsoft Visio 2013 | application/vnd.visio | VSDX, VSTX, | adPRESENTATION |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|----------------------------------|--------|----------|-------------------------------------|--|------------------|-----------------|
| | | | | | VSSX | |
| MS_Visio_2013_Macro_Fmt | 468 | 415 | Microsoft Visio 2013 macro | application/vnd.visio | VSDM, VSTM, VSSM | adPRESENTATION |
| ICHITARO_Compr_Fmt | 469 | 417 | ICHITARO Compressed format | application/x-js-taro | JTDC | adWORDPROCESSOR |
| IWWP13_Fmt | 470 | 418 | Apple iWork 2013 Pages format | | IWA, PAGES | adWORDPROCESSOR |
| IWSS13_Fmt | 471 | 419 | Apple iWork 2013 Numbers format | | IWA, NUMBERS | adSPREADSHEET |
| IWPG13_Fmt | 472 | 420 | Apple iWork 2013 Keynote format | | IWA, KEY | adPRESENTATION |
| XZ_Fmt | 473 | 421 | XZ archive format | application/x-xz | XZ | adENCAPSULATION |
| Sony_WAVE64_Fmt | 474 | 422 | Sony Wave64 format | audio/wav64 | W64 | adSOUND |
| Conifer_WAVPACK_Fmt | 475 | 423 | Conifer Wavpack format | audio/x-wavpack | WV | adSOUND |
| Xiph_OGG_VORBIS_Fmt | 476 | 424 | Xiph Ogg Vorbis format | audio/ogg | OGG | adSOUND |
| MS_Visio_2013_Stencil_Fmt | 477 | 415 | MS Visio 2013 stencil format | application/vnd.visio | VSSX | adPRESENTATION |
| MS_Visio_2013_Stencil_Macro_Fmt | 478 | 415 | MS Visio 2013 stencil Macro format | application/vnd.visio | VSSM | adPRESENTATION |
| MS_Visio_2013_Template_Fmt | 479 | 415 | MS Visio 2013 template format | application/vnd.visio | VSTX | adPRESENTATION |
| MS_Visio_2013_Template_Macro_Fmt | 480 | 415 | MS Visio 2013 template Macro format | application/vnd.visio | VSTM | adPRESENTATION |
| Borland_Reflex_2_Fmt | 481 | 425 | Borland Reflex 2 format | | R2D | adDATABASE |
| PKCS_12_Fmt | 482 | 426 | PKCS #12 (p12) format | application/x-pkcs12 | P12, PFX | adWORDPROCESSOR |
| B1_Fmt | 483 | 427 | B1 format | application/x-b1 | B1 | adENCAPSULATION |
| ISO_IEC_MPEG_4_Fmt | 484 | 428 | ISO/IEC MPEG-4 (ISO 14496) format | video/mp4 | MP4 | adMOVIE |
| RAR5_Fmt | 485 | 429 | RAR5 Format | application/x-rar-compressed | RAR | adENCAPSULATION |
| Unigraphics_NX_Fmt | 486 | 362 | Unigraphics (UG) NX CAD Format | | PRT | adVECTORGRAPHIC |
| PTC_Creo_Fmt | 487 | 430 | PTC Creo CAD Format | | ASM, PRT | adVECTORGRAPHIC |
| KML_Fmt | 488 | 431 | Keyhole Markup Language | application/vnd.google-earth.kml+xml | KML | adWORDPROCESSOR |
| KMZ_Fmt | 489 | 432 | Zipped Keyhole Markup Language | application/vnd.google-earth.kmz | KMZ | adWORDPROCESSOR |
| WML_Fmt | 490 | 433 | Wireless Markup Language | text/vnd.wap.wml | WML | adWORDPROCESSOR |
| ODF_Formula_Fmt | 491 | 434 | ODF Formula | application/vnd.oasis.opendocument.formula | ODF | adWORDPROCESSOR |
| SO_Text_Fmt | 492 | 435 | Star Office 4,5 Writer Text | application/vnd.stardivision.writer | SDW, SGL, | adWORDPROCESSOR |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|---------------------|--------|----------|--|-----------------------------------|--------------|----------------|
| | | | | | VOR | |
| SO_Spreadsheet_Fmt | 493 | 436 | Star Office 4,5 Calc Spreadsheet | application/vnd.stardivision.calc | SDC | adSPREADSHEET |
| SO_Presentation_Fmt | 494 | 437 | Star Office 4,5 Impress Presentation | application/vnd.stardivision.draw | SDD, SDA | adPRESENTATION |
| SO_Math_Fmt | 495 | 438 | Star Office 4,5 Math | application/vnd.stardivision.math | SMF | adMISC |
| STEP_Fmt | 496 | 439 | ISO 10303-21 STEP format | | | adMISC |
| STL_Fmt | 497 | 364 | 3D Systems STL ASCII format | | | adMISC |
| AppleScript_Fmt | 498 | 440 | AppleScript Source Code ³ | text/x-applescript | APPLESCRIPT | adSOURCECODE |
| Assembly_Fmt | 499 | 441 | Assembly Code ³ | text/x-assembly | | adSOURCECODE |
| C_Fmt | 500 | 442 | C Source Code ³ | text/x-c | C, H | adSOURCECODE |
| Csharp_Fmt | 501 | 443 | C# Source Code ³ | text/x-csharp | CS | adSOURCECODE |
| CPlusPlus_Fmt | 502 | 444 | C++ Source Code ³ | text/x-c++ | CPP, HPP | adSOURCECODE |
| Css_Fmt | 503 | 445 | Cascading Style Sheet ³ | text/css | CSS | adSOURCECODE |
| Clojure_Fmt | 504 | 446 | Clojure Source Code ³ | text/x-clojure | CLJ, CL2 | adSOURCECODE |
| CoffeeScript_Fmt | 505 | 447 | CoffeeScript Source Code ³ | text/x-coffeescript | COFFEE, CAKE | adSOURCECODE |
| Lisp_Fmt | 506 | 448 | Common Lisp Source Code ³ | text/x-common-lisp | EL | adSOURCECODE |
| Dockerfile_Fmt | 507 | 449 | Dockerfile ³ | text/x-dockerfile | | adSOURCECODE |
| Eiffel_Fmt | 508 | 450 | Eiffel Source Code ³ | text/x-eiffel | E | adSOURCECODE |
| Erlang_Fmt | 509 | 451 | Erlang Source Code ³ | text/x-erlang | ERL, ES | adSOURCECODE |
| Fsharp_Fmt | 510 | 452 | F# Source Code ³ | text/x-fsharp | FS | adSOURCECODE |
| Fortran_Fmt | 511 | 453 | Fortran Source Code ³ | text/x-fortran | F | adSOURCECODE |
| Go_Fmt | 512 | 454 | Go Source Code ³ | text/x-go | GO | adSOURCECODE |
| Groovy_Fmt | 513 | 455 | Groovy Source Code ³ | text/x-groovy | GRT, GVV | adSOURCECODE |
| Haskell_Fmt | 514 | 456 | Haskell Source Code ³ | text/x-haskell | HS | adSOURCECODE |
| Ini_Fmt | 515 | 457 | Initialization (INI) file ³ | text/x-ini | | adSOURCECODE |
| Java_Fmt | 516 | 458 | Java Source Code ³ | text/x-java-source | JAVA | adSOURCECODE |
| Javascript_Fmt | 517 | 459 | Javascript Source Code ³ | text/javascript | JS | adSOURCECODE |
| Lua_Fmt | 518 | 460 | Lua Source Code ³ | text/x-lua | LUA | adSOURCECODE |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|-----------------|--------|----------|--|---------------------|-------------|-----------------|
| Makefile_Fmt | 519 | 461 | Makefile ³ | text/x-makefile | MAKE | adSOURCECODE |
| Mathematica_Fmt | 520 | 462 | Wolfram Mathematica Source Code ³ | text/x-mathematica | M | adSOURCECODE |
| ObjC_Fmt | 521 | 464 | Objective-C Source Code ³ | text/x-objc | | adSOURCECODE |
| ObjCpp_Fmt | 522 | 465 | Objective-C++ Source Code ³ | text/x-objectivec++ | | adSOURCECODE |
| ObjJ_Fmt | 523 | 466 | Objective-J Source Code ³ | text/x-objectivej | J | adSOURCECODE |
| PHP_Fmt | 524 | 467 | PHP Source Code ³ | text/x-php | PHP | adSOURCECODE |
| PLSQL_Fmt | 525 | 468 | PLSQL Source Code ³ | text/x-plsql | | adSOURCECODE |
| Pascal_Fmt | 526 | 469 | Pascal Source Code ³ | text/x-pascal | PASCAL | adSOURCECODE |
| Perl_Fmt | 527 | 470 | Perl Source Code ³ | text/x-perl | PL | adSOURCECODE |
| Powershell_Fmt | 528 | 471 | PowerShell Source Code ³ | text/x-powershell | PS1 | adSOURCECODE |
| Prolog_Fmt | 529 | 472 | Prolog Source Code ³ | text/x-prolog | PRO, PROLOG | adSOURCECODE |
| Puppet_Fmt | 530 | 473 | Puppet Source Code ³ | text/x-puppet | PP | adSOURCECODE |
| Python_Fmt | 531 | 474 | Python Source Code ³ | text/x-python | PY | adSOURCECODE |
| R_Fmt | 532 | 475 | R Source Code ³ | text/x-rsrc | R | adSOURCECODE |
| Ruby_Fmt | 533 | 476 | Ruby Source Code ³ | text/x-ruby | RB | adSOURCECODE |
| Rust_Fmt | 534 | 477 | Rust Source Code ³ | text/x-rust | RS | adSOURCECODE |
| Scala_Fmt | 535 | 478 | Scala Source Code ³ | text/x-scala | SC | adSOURCECODE |
| Shell_Fmt | 536 | 479 | Shell Script ³ | application/x-sh | SH | adSOURCECODE |
| Smalltalk_Fmt | 537 | 480 | Smalltalk Source Code ³ | text/x-stsrc | ST | adSOURCECODE |
| ML_Fmt | 538 | 481 | Standard ML Source Code ³ | text/x-ml | ML | adSOURCECODE |
| Swift_Fmt | 539 | 482 | Swift Source Code ³ | text/x-swift | SWIFT | adSOURCECODE |
| Tcl_Fmt | 540 | 483 | Tool Command Language (Tcl) Source Code ³ | text/x-tcl | TM | adSOURCECODE |
| Tex_Fmt | 541 | 484 | TeX Typesetting File ³ | application/x-tex | | adSOURCECODE |
| TypeScript_Fmt | 542 | 485 | TypeScript Source Code ³ | text/x-typescript | TS | adSOURCECODE |
| Verilog_Fmt | 543 | 486 | Verilog Source Code ³ | text/x-verilog | V | adSOURCECODE |
| YAML_Fmt | 544 | 487 | YAML File ³ | text/x-yaml | YML | adSOURCECODE |
| Wiki_Fmt | 545 | 488 | MediaWiki File | text/x-mediawiki | | adWORDPROCESSOR |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|---------------------------|--------|----------|------------------------------------|-------------------------------|---------------|-----------------|
| MS_Word_2007_Flat_XML_Fmt | 546 | 301 | Microsoft Word 2007 XML - Flat xml | text/xml | XML | adWORDPROCESSOR |
| Matroska_Fmt | 547 | 489 | Matroska video File | video/x-matroska | MKV | adMOVIE |
| SVG_Fmt | 548 | 490 | Scalable Vector Graphics image | image/svg+xml | SVG | adVECTORGRAPHIC |
| Shapefile_Fmt | 549 | 491 | Shapefile | application/x-shapefile | SHP, SHX | adGIS |
| Flash_Video_Fmt | 550 | 492 | Flash video File | video/x-flv | FLV | adMOVIE |
| Embedded_OpenType_Fmt | 551 | 493 | Embedded OpenType font | application/vnd.ms-fontobject | EOT | adFONT |
| Web_Open_Font_Fmt | 552 | 494 | Web Open Font Format | font/woff | WOFF, WOFF2 | adFONT |
| OpenType_Fmt | 553 | 495 | OpenType Font | font/otf | OTF | adFONT |
| MNG_Fmt | 554 | 496 | Multiple-image Network Graphics | video/x-mng | MNG | adANIMATION |
| JNG_Fmt | 555 | 497 | JPEG Network Graphics | image/x-jng | JNG | adRASTERIMAGE |
| AppleScript_Binary_Fmt | 556 | 498 | AppleScript Binary Source Code | | SCPT | adSOURCECODE |
| Maya_Binary_Fmt | 557 | 499 | Autodesk Maya binary file | | MB | adCAD |
| Jupiter_Tesselation_Fmt | 558 | 363 | UGS Jupiter Tesselation file | | JT | adCAD |
| OGV_Fmt | 559 | 500 | Ogg Theora Video format | video/ogg | OGV | adMOVIE |
| OGG_Container_Fmt | 560 | 501 | General Ogg Container format | application/ogg | OGG | adMISC |
| GNU_Message_Catalog_Fmt | 561 | 502 | GNU Message Catalog format | | MO | adMISC |
| Windows_Shortcut_Fmt | 562 | 503 | Windows shortcut file | application/x-ms-shortcut | LNK | adMISC |
| Apple_Typedstream_Fmt | 563 | 504 | Apple/NeXT typedstream data format | | | adMISC |
| XCF_Fmt | 564 | 505 | GIMP XCF image | image/x-xcf | XCF | adRASTERIMAGE |
| PaintShop_Pro_Fmt | 565 | 506 | PaintShop Pro image | | PSP, PSPIMAGE | adRASTERIMAGE |
| SQLite_Database_Fmt | 566 | 507 | SQLite database format | application/x-sqlite3 | QHC | adDATABASE |
| MySQL_Table_Fmt | 567 | 508 | MySQL table definition file | | FRM | adDATABASE |
| Microsoft_Program_DB_Fmt | 568 | 509 | Microsoft Program Database format | | PDB | adDATABASE |
| OpenEXR_Fmt | 569 | 510 | OpenEXR image format | | EXR | adRASTERIMAGE |
| XMV_Fmt | 570 | 511 | 4X Movie File | | | adMOVIE |
| AMV_Fmt | 571 | 512 | AMV video file | | AMV | adMOVIE |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|------------------------|--------|----------|--|----------------------------|-----------|-----------------|
| NIFF_Fmt | 572 | 513 | Notation Interchange File Format | | NIF | adSOUND |
| CuBase_Fmt | 573 | 514 | Steinberg CuBase file | | | adSOUND |
| SoundFont_Fmt | 574 | 515 | SoundFont file | | | adSOUND |
| WebP_Fmt | 575 | 516 | WebP image | image/webp | WEBP | adRASTERIMAGE |
| ICC_Fmt | 576 | 517 | International Color Consortium files | application/vnd.iccprofile | ICC, ICM | adMISC |
| PCF_Fmt | 577 | 518 | X11 Portable Compiled Font file | application/x-font-pcf | PCF | adFONT |
| WebM_Fmt | 578 | 519 | WebM video file | video/webm | WEBM | adMOVIE |
| AMFF_Fmt | 579 | 520 | Amiga Metafile | | AMF | adVECTORGRAPHIC |
| ANBM_Fmt | 580 | 521 | IFF Animated Bitmap | | | adRASTERIMAGE |
| ANIM_Fmt | 581 | 522 | IFF Amiga animated raster graphics format | | | adRASTERIMAGE |
| DEEP_Fmt | 582 | 523 | IFF-DEEP TVPaint image | | DEEP | adRASTERIMAGE |
| FAXX_Fmt | 583 | 524 | IFF-FAXX Facsimile image | | | adRASTERIMAGE |
| ICON_Fmt | 584 | 525 | IFF Glow Icon image | | | adRASTERIMAGE |
| ILBM_Fmt | 585 | 526 | Interleaved BitMap image | | IFF | adRASTERIMAGE |
| LWOB_Fmt | 586 | 527 | LightWave Object format | | LWOB | adMISC |
| MAUD_Fmt | 587 | 528 | IFF-MAUD MacroSystem audio format | | | adSOUND |
| PBM_Fmt | 588 | 529 | IFF Planar BitMap | | | adRASTERIMAGE |
| TDDD_Fmt | 589 | 530 | IFF TDDD and Imagine Object animation format | | TDD | adRASTERIMAGE |
| DjVu_Fmt | 590 | 531 | AT&T DjVu format | image/vnd.djvu | DJVU | adWORDPROCESSOR |
| InDesign_Fmt | 591 | 532 | Adobe InDesign document | application/x-indesign | | adDESKTOPPUBLSH |
| Calamus_Fmt | 592 | 533 | Calamus Desktop Publishing | | | adDESKTOPPUBLSH |
| Adaptive_MultiRate_Fmt | 593 | 534 | Adaptive Multi-Rate audio format | audio/amr | AMR | adSOUND |
| FLAC_Fmt | 594 | 535 | Free Lossless Audio Codec format | audio/flac | FLAC | adSOUND |
| Ogg_FLAC_Fmt | 595 | 536 | Ogg Container FLAC audio format | | OGG | adSOUND |
| SAS7BDAT_Fmt | 596 | 537 | SAS7BDAT database storage format | | SAS7BDAT | adDATABASE |
| Design_Web_Format_Fmt | 597 | 538 | Autodesk Design Web Format | model/vnd.dwf | DWF | adCAD |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|---------------------------------|--------|----------|--|---------------------|-----------|-----------------|
| Adobe_Flash_Audio_Book_Fmt | 598 | 539 | Adobe Flash Player audio book | audio/mp4 | F4B | adSOUND |
| Adobe_Flash_Audio_Fmt | 599 | 540 | Adobe Flash Player audio | audio/mp4 | F4A | adSOUND |
| Adobe_Flash_Protected_Video_Fmt | 600 | 541 | Adobe Flash Player protected video | video/mp4 | F4P | adMOVIE |
| Adobe_Flash_Video_Fmt | 601 | 542 | Adobe Flash Player video | video/x-f4v | F4V | adMOVIE |
| Audible_Audiobook_Fmt | 602 | 543 | Audible Enhanced Audiobook | | AAX | adSOUND |
| Canon_Camera_Fmt | 603 | 544 | Canon Digital Camera image | | | adRASTERIMAGE |
| Canon_Raw_Fmt | 604 | 545 | Canon Raw image | | CR3 | adRASTERIMAGE |
| Casio_Camera_Fmt | 605 | 546 | Casio Digital Camera image | | | adRASTERIMAGE |
| Convergent_Design_Fmt | 606 | 547 | Convergent Design file | | | adRASTERIMAGE |
| DMB_MAF_Audio_Fmt | 607 | 548 | DMB MAF audio | | | adSOUND |
| DMB_MAF_Video_Fmt | 608 | 549 | DMB MAF video | | | adMOVIE |
| DMP_Content_Fmt | 609 | 550 | Digital Media Project Content Format | | | adMISC |
| DVB_Fmt | 610 | 551 | Digital Video Broadcast format | video/vnd.dvb.file | DVB | adMOVIE |
| Dirac_Wavelet_Compression_Fmt | 611 | 552 | ISO-BMFF Dirac Wavelet compression | | | adMISC |
| HEICS_Image_Sequence_Fmt | 612 | 553 | High Efficiency Image Format HEVC image sequence | image/heic-sequence | HEICS | adRASTERIMAGE |
| HEIC_Image_Fmt | 613 | 554 | High Efficiency Image Format HEVC image | image/heic | HEIC | adRASTERIMAGE |
| HEIFS_Image_Sequence_Fmt | 614 | 555 | High Efficiency Image Format image sequence | image/heif-sequence | HEIFS | adRASTERIMAGE |
| HEIF_Image_Fmt | 615 | 556 | High Efficiency Image Format image | image/heif | HEIF | adRASTERIMAGE |
| ISMACryp_Fmt | 616 | 557 | ISMACryp 2.0 Encrypted format | | | adENCAPSULATION |
| ISO_3GPP2_Fmt | 617 | 558 | 3GPP2 video file | video/3gpp2 | 3G2 | adMOVIE |
| ISO_3GPP_Fmt | 618 | 559 | 3GPP video file | video/3gpp | 3GP | adMOVIE |
| ISO_JPEG2000_JP2_Fmt | 619 | 560 | ISO-BMFF JPEG 2000 image | image/jp2 | JP2 | adRASTERIMAGE |
| ISO_JPEG2000_JPM_Fmt | 620 | 561 | ISO-BMFF JPEG 2000 compound image | image/jpm | JPM | adRASTERIMAGE |
| ISO_JPEG2000_JPX_Fmt | 621 | 562 | ISO-BMFF JPEG 2000 with extensions | image/jpx | JPX | adRASTERIMAGE |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|----------------------------------|--------|----------|--|-----------------|-----------|---------------|
| ISO_QuickTime_Fmt | 622 | 563 | Apple ISO-BMFF QuickTime video | video/quicktime | QT, MOV | adMOVIE |
| KDDI_Video_Fmt | 623 | 564 | KDDI Video file | video/3gpp2 | | adMOVIE |
| MAF_Photo_Player_Fmt | 624 | 565 | MAF Photo Player | | | adMISC |
| MPEG4_AVC_Fmt | 625 | 566 | ISO-BMFF MPEG-4 with AVC extension | video/mp4 | | adMOVIE |
| MPEG4_M4A_Fmt | 626 | 567 | Apple MPEG-4 Part 14 audio | audio/x-m4a | M4A | adSOUND |
| MPEG4_M4B_Fmt | 627 | 568 | Apple MPEG-4 Part 14 audio book | audio/mp4 | M4B | adSOUND |
| MPEG4_M4P_Fmt | 628 | 569 | Apple MPEG-4 Part 14 protected audio | audio/mp4 | M4P | adSOUND |
| MPEG4_M4V_Fmt | 629 | 570 | Apple MPEG-4 Part 14 video | video/x-m4v | M4V | adMOVIE |
| MPEG4_Sony_PSP_Fmt | 630 | 571 | Sony PSP MPEG-4 | audio/mp4 | MP4 | adSOUND |
| MPEG_21_Fmt | 631 | 572 | MPEG-21 | audio/mp4 | | adMISC |
| Mobile_QuickTime_Fmt | 632 | 573 | Mobile QuickTime video | video/quicktime | MQV | adMOVIE |
| Motion_JPEG_2000_Fmt | 633 | 574 | Motion JPEG 2000 | video/mj2 | MJ2, MJP2 | adMOVIE |
| NTT_MPEG4_Fmt | 634 | 575 | NTT MPEG-4 | video/mp4 | | adMOVIE |
| Nero_MPEG4_AVC_Profile | 635 | 576 | Nero MPEG-4 profile with AVC extension | video/mp4 | | adMOVIE |
| Nero_MPEG4_Audio_Fmt | 636 | 577 | Nero AAC audio | audio/mp4 | | adSOUND |
| Nero_MPEG4_Profile | 637 | 578 | Nero MPEG-4 profile | video/mp4 | | adMOVIE |
| OMA_DRM_Fmt | 638 | 579 | OMA DRM Format | | | adMISC |
| Panasonic_Camera_Fmt | 639 | 580 | Panasonic Digital Camera image | | | adRASTERIMAGE |
| Ross_Video_Fmt | 640 | 581 | Ross video | | | adMOVIE |
| SDA_Video_Fmt | 641 | 582 | SDA SD Memory Card video | | | adMOVIE |
| Samsung_Stereoscopic_Fmt | 642 | 583 | Samsung stereoscopic stream | | | adMISC |
| Sony_XAVC_Fmt | 643 | 584 | Sony XAVC video | | | adMOVIE |
| JPEG_2000_PGX_Fmt | 644 | 585 | JPEG 2000 PGX Verification Model image | | PGX | adRASTERIMAGE |
| Apple_Desktop_Services_Store_Fmt | 645 | 586 | Apple Desktop Services Store file | | DS_Store | adMISC |
| Core_Audio_Fmt | 646 | 587 | Apple Core Audio Format | audio/x-caf | CAF | adSOUND |
| VICAR_Fmt | 647 | 588 | VICAR image format | | IMG | adRASTERIMAGE |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|---------------------------|--------|----------|---|----------------------|-----------|---------------|
| FITS_Fmt | 648 | 589 | Flexible Image Transport System FITS image | image/fits | FIT | adRASTERIMAGE |
| DIF_Fmt | 649 | 590 | Digital Interface Format (DIF) DV video | | DV | adMOVIE |
| MPEG_Transport_Stream_Fmt | 650 | 591 | MPEG Transport Stream data | video/MP2T | TS | adMISC |
| MPEG_Sequence_Fmt | 651 | 592 | MPEG Sequence format | video/mpeg | | adMISC |
| Ogg_OGM_Fmt | 652 | 593 | Ogg OGM video format | video/ogg | OGM | adMOVIE |
| Ogg_Speex_Fmt | 653 | 594 | Ogg Speex audio format | audio/ogg | SPX | adSOUND |
| Ogg_Opus_Fmt | 654 | 595 | Ogg Opus audio format | audio/ogg | OGG | adSOUND |
| Musepack_Audio_Fmt | 655 | 596 | Musepack audio format | audio/x-musepack | MPC | adSOUND |
| ART_Image_Fmt | 656 | 597 | ART image format | | ART | adRASTERIMAGE |
| Vivo_Fmt | 657 | 598 | Vivo audio-video format | video/vnd.vivo | VIV | adMOVIE |
| QCP_Fmt | 658 | 599 | Qualcomm QCP audio | audio/qcelp | QCP | adSOUND |
| CSP_Codec_Fmt | 659 | 600 | Creative Signal Processor codec | | CSP | adMISC |
| TwinVQ_Fmt | 660 | 601 | NTT TwinVQ audio format | | VQF | adSOUND |
| Interplay_MVE_Fmt | 661 | 602 | Interplay MVE video format | | MVE | adMOVIE |
| IRIX_Moviemaker_Fmt | 662 | 603 | IRIX Silicon Graphics moviemaker video file | video/x-sgi-movie | MV, MOVIE | adMOVIE |
| Sega_FILM_Fmt | 663 | 604 | Sega FILM video format | | CPK, CAK | adMOVIE |
| SMAF_Fmt | 664 | 605 | Synthetic music Mobile Application Format | application/vnd.smaf | MMF | adSOUND |
| NIST_SPHERE_Fmt | 665 | 606 | NIST SPeech HEader REsources format | | NIST | adSOUND |
| Chinese_AVS_Fmt | 666 | 607 | Chinese AVS video format | | | adMOVIE |
| VQA_Fmt | 667 | 608 | Westwood Studios Vector Quantized Animation video file | | VQA | adANIMATION |
| YAFA_Fmt | 668 | 609 | Wildfire YAFA animation | | YAFA | adANIMATION |
| Origin_MVE_Fmt | 669 | 610 | Origin Wing Commander III MVE movie format | | MVE | adMOVIE |
| BBC_Dirac_Fmt | 670 | 611 | BBC Dirac video format | video/x-dirac | DRC | adMOVIE |
| Maya_ASCII_Fmt | 671 | 612 | Autodesk Maya ASCII file format | | MA | adCAD |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|------------------------------|--------|----------|---|-----------------------------------|-----------|-----------------|
| RenderMan_Fmt | 672 | 613 | Pixar RenderMan Interface Bytestream file | | RIB | adVECTORGRAPHIC |
| NOFF_Binary_Fmt | 673 | 614 | NOFF 3D Object File Format | | NOFF | adVECTORGRAPHIC |
| VTK_ASCII_Fmt | 674 | 615 | Visualization Toolkit VTK ASCII format | | VTK | adVECTORGRAPHIC |
| VTK_Binary_Fmt | 675 | 616 | Visualization Toolkit VTK Binary format | | VTK | adVECTORGRAPHIC |
| Wolfram_CDF_Fmt | 676 | 617 | Wolfram Mathematica Computable Document Format | application/cdf | CDF | adMISC |
| Wolfram_Notebook_Fmt | 677 | 618 | Wolfram Mathematica Notebook Format | | NB | adMISC |
| HDF4_Fmt | 678 | 619 | Hierarchical Data Format HDF4 | application/x-hdf | HDF, H4 | adMISC |
| HDF5_Fmt | 679 | 620 | Hierarchical Data Format HDF5 | application/x-hdf | HDF, H5 | adMISC |
| ARMovie_Fmt | 680 | 621 | Acorn RISC ARMovie video format | | RPL | adMOVIE |
| Windows_TV_DVR_Fmt | 681 | 622 | Windows Television DVR format | | WTV | adMOVIE |
| InstallShield_Z_Fmt | 682 | 623 | InstallShield Z archive format | application/x-compress | Z | adENCAPSULATION |
| MS_DirectDraw_Surface_Fmt | 683 | 624 | Microsoft DirectDraw Surface container format | | DDS | adENCAPSULATION |
| Bink_Fmt | 684 | 625 | Bink audio-video container format | | BIK, BK2 | adMOVIE |
| LZMA_Fmt | 685 | 626 | LZMA compressed data format | application/x-lzma | LZMA | adENCAPSULATION |
| True_Audio_Fmt | 686 | 627 | True Audio format | audio/x-tta | TTA | adSOUND |
| Keepass_Fmt | 687 | 628 | Keepass Password file | | KDB, KDBX | adMISC |
| RPM_Fmt | 688 | 629 | RPM Package Manager file | application/x-rpm | RPM | adENCAPSULATION |
| Printer_Font_Metrics_Fmt | 689 | 630 | Adobe Printer Font Metrics format | application/x-font-printer-metric | PFM | adFONT |
| Adobe_Font_Metrics_Fmt | 690 | 631 | Adobe Font Metrics ASCII format | application/x-font-adobe-metric | AFM | adFONT |
| Printer_Font_ASCII_Fmt | 691 | 632 | Adobe Printer Font ASCII format | application/x-font-type1 | PFA | adFONT |
| Netware_Loadable_Module_Fmt | 692 | 633 | Netware Loadable Module format | | NLM | adMISC |
| TCPdump_pcap_Fmt | 693 | 634 | TCPdump packet stream capture savefile format | application/vnd.tcpdump.pcap | PCAP | adMISC |
| Multiple_Master_Font_Fmt | 694 | 635 | Adobe Multiple master font format | | MMM | adFONT |
| TrueType_Font_Collection_Fmt | 695 | 636 | TrueType font collection format | application/x-font-ttf | TTC | adFONT |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|------------------------------------|--------|----------|---|-----------------------------------|--------------------|------------------|
| Shapefile_Spatial_Index_Fmt | 696 | 637 | Shapefile binary spatial index format | application/x-shapefile | SBX, SBN | adGIS |
| Java_Key_Store_Fmt | 697 | 638 | Java Key Store format | application/x-java-keystore | KS | adMISC |
| Java_JCE_Key_Store_Fmt | 698 | 639 | Java JCE Key Store format | application/x-java-jce-keystore | | adMISC |
| Quark_Xpress_Intel_Fmt | 699 | 640 | QuarkXPress Intel format | application/vnd.quark.quarkxpress | QXB | adDESKTOPPUBLISH |
| Windows_Imaging_Fmt | 700 | 641 | Microsoft Windows Imaging Format WIM | | WIM | adMISC |
| VMware_Virtual_Disk_Fmt | 701 | 642 | VMware Virtual Disk Format 5.0 | application/x-vmdk | VMDK | adMISC |
| XPCConnect_Typelib_Fmt | 702 | 643 | XPCConnect Typelib Format | | XPT | adMISC |
| MS_DOS_Compression_Fmt | 703 | 644 | Microsoft MS-DOS installation 'Quantum' compression | | EX_ | adENCAPSULATION |
| DLS_Fmt | 704 | 645 | DLS Downloadable Sounds format | | DLS | adSOUND |
| MS_Windows_Registry_Fmt | 705 | 646 | Microsoft Windows Registry format | | | adMISC |
| Microsoft_Help_2_Fmt | 706 | 647 | Microsoft Help 2.0 format | | HXD, HXW, HXH | adENCAPSULATION |
| Qt_Translation_Fmt | 707 | 648 | Qt binary translation file format | | QM | adMISC |
| PEM_SSL_Certificate_Fmt | 708 | 649 | PEM-encoded SSL certificate | application/pkix-cert | CRT, PEM, CER, KEY | adENCAPSULATION |
| PostScript_Printer_Description_Fmt | 709 | 650 | Adobe PostScript Printer Description file | application/vnd.cups-ppd | PPD | adMISC |
| Speedo_Font_Fmt | 710 | 651 | Speedo Font format | | SPD | adFONT |
| InstallShield_Cabinet_Fmt | 711 | 652 | InstallShield Cabinet Archive format | | CAB, HDR | adENCAPSULATION |
| InstallShield_Uninstall_Fmt | 712 | 653 | InstallShield Uninstall format | | ISU | adENCAPSULATION |
| MS_OEDBX_Folder_Fmt | 713 | 654 | Outlook Express DBX folder database format | | DBX | adENCAPSULATION |
| LabVIEW_Fmt | 714 | 655 | National Instruments LabVIEW file format | | VI | adMISC |
| SAP_Archive_SAR_Fmt | 715 | 656 | SAP compression archive SAR format | | SAR | adENCAPSULATION |
| Netscape_Address_Book_Fmt | 716 | 657 | Netscape Address Book format | | NAB | adMISC |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|-----------------------------------|--------|----------|---|------------------------------|-----------|-----------------|
| Universal_3D_Fmt | 717 | 658 | Universal 3D file format | | U3D | adVECTORGRAPHIC |
| Open_Inventor_ASCII_Fmt | 718 | 659 | Open Inventor ASCII format | | IV | adVECTORGRAPHIC |
| Open_Inventor_Binary_Fmt | 719 | 660 | Open Inventor Binary format | | IV | adVECTORGRAPHIC |
| X_Window_Dump_Fmt | 720 | 661 | X Window Dump image | image/x-xwindowdump | XWD | adRASTERIMAGE |
| Git_Packfile_Fmt | 721 | 662 | Git Packfile format | | PACK | adENCAPSULATION |
| Xara_Xar_Fmt | 722 | 663 | Xara X Xar image format | application/vnd.xara | XAR | adVECTORGRAPHIC |
| Internet_Archive_ARC_Fmt | 723 | 664 | Internet Archive ARC format | application/x-ia-arc | ARC | adENCAPSULATION |
| Applix_Builder_Fmt | 724 | 665 | Applix Builder format | | AB | adMISC |
| Applix_Bitmap_Fmt | 725 | 666 | Applix Bitmap image format | | IM | adRASTERIMAGE |
| PEM_RSA_Private_Key_Fmt | 726 | 667 | PEM-encoded RSA private key | | PEM | adENCAPSULATION |
| MIFF_Fmt | 727 | 668 | Magick Image File Format | | MIFF | adRASTERIMAGE |
| Subversion_Dump_Fmt | 728 | 669 | Subversion Dump format | | | adENCAPSULATION |
| Virtual_Hard_Disk_Fmt | 729 | 670 | Microsoft Virtual Hard Disk format | application/x-vhd | VHD | adENCAPSULATION |
| Direct_Access_Archive_Fmt | 730 | 671 | PowerISO Direct Access Archive format | | DAA | adENCAPSULATION |
| Debian_Binary_Fmt | 731 | 672 | Debian binary package format | application/x-debian-package | DEB | adENCAPSULATION |
| XUL_Fastload_Fmt | 732 | 673 | Mozilla XUL Fastload format | | MFL | adMISC |
| Nastran_OP2_Fmt | 733 | 674 | Nastran OP2 format | | OP2 | adCAD |
| Binary_Logging_Fmt | 734 | 675 | CAD Binary Logging Format | | BLF | adCAD |
| Measurement_Data_Fmt | 735 | 676 | CAD Measurement Data Format | | MDF | adCAD |
| Abaqus_ODB_Fmt | 736 | 677 | Abaqus ODB Format | | ODB | adCAD |
| Open_Diagnostic_Data_Exchange_Fmt | 737 | 678 | Vector Open Diagnostic Data Exchange format | | ODX | adCAD |
| Vector_ASCII_Fmt | 738 | 679 | Vector CAD ASCII ASC format | | ASC | adCAD |
| LSDYNA_State_Database_Fmt | 739 | 680 | LS-DYNA State Database format | | | adCAD |
| LSDYNA_Binary_Output_Fmt | 740 | 681 | LS-DYNA binary output (binout) format | | | adCAD |
| MS_Power_BI_Fmt | 741 | 682 | Microsoft Power BI Desktop format | | PBIX | adANALYTICS |
| Tableau_Workbook_Fmt | 742 | 683 | Tableau Workbook format | | TWB | adANALYTICS |
| Tableau_Packaged_Workbook_Fmt | 743 | 684 | Tableau Packaged Workbook | | TWBX | adANALYTICS |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|----------------------------------|--------|----------|---------------------------------------|------------------|--------------------|--------------|
| | | | format | | | |
| Tableau_Extract_Fmt | 744 | 685 | Tableau Extract format | | TDE | adANALYTICS |
| Tableau_Data_Source_Fmt | 745 | 686 | Tableau Data Source format | | TDS | adANALYTICS |
| Tableau_Packaged_Data_Source_Fmt | 746 | 687 | Tableau Packaged Data Source format | | TDSX | adANALYTICS |
| Tableau_Preferences_Fmt | 747 | 688 | Tableau Preferences format | | TPS | adANALYTICS |
| Tableau_Map_Source_Fmt | 748 | 689 | Tableau Map Source format | | TMS | adANALYTICS |
| ABAP_Fmt | 749 | 690 | ABAP Source Code ⁴ | text/x-abap | ABAP | adSOURCECODE |
| AMPL_Fmt | 750 | 691 | AMPL Source Code ⁴ | | AMPL | adSOURCECODE |
| APL_Fmt | 751 | 692 | APL Source Code ⁴ | | APL | adSOURCECODE |
| ASN1_Fmt | 752 | 693 | ASN.1 Source Code ⁴ | | ASN | adSOURCECODE |
| ATS_Fmt | 753 | 694 | ATS Source Code ⁴ | | | adSOURCECODE |
| Agda_Fmt | 754 | 695 | Agda Source Code ⁴ | text/x-agda | AGDA | adSOURCECODE |
| Alloy_Fmt | 755 | 696 | Alloy Source Code ⁴ | text/x-alloy | ALS | adSOURCECODE |
| Apex_Fmt | 756 | 697 | Apex Source Code ⁴ | | CLS | adSOURCECODE |
| Arduino_Fmt | 757 | 698 | Arduino Source Code ⁴ | text/x-arduino | INO | adSOURCECODE |
| AsciiDoc_Fmt | 758 | 699 | AsciiDoc Source Code ⁴ | text/x-asciidoc | ASC | adSOURCECODE |
| AspectJ_Fmt | 759 | 700 | AspectJ Source Code ⁴ | text/x-aspectj | AJ | adSOURCECODE |
| Awk_Fmt | 760 | 701 | Awk Source Code ⁴ | text/x-awk | AWK | adSOURCECODE |
| BlitzMax_Fmt | 761 | 702 | BlitzMax Source Code ⁴ | text/x-bmx | BMX | adSOURCECODE |
| Bluespec_Fmt | 762 | 703 | Bluespec Source Code ⁴ | | BSV | adSOURCECODE |
| Brainfuck_Fmt | 763 | 704 | Brainfuck Source Code ⁴ | text/x-brainfuck | B, BF | adSOURCECODE |
| Brightscript_Fmt | 764 | 705 | Brightscript Source Code ⁴ | | BRS | adSOURCECODE |
| CLIPS_Fmt | 765 | 706 | CLIPS Source Code ⁴ | | CLP | adSOURCECODE |
| CMake_Fmt | 766 | 707 | CMake Source Code ⁴ | text/x-cmake | CMAKE | adSOURCECODE |
| COBOL_Fmt | 767 | 708 | COBOL Source Code ⁴ | text/x-cobol | CBL, CCP, COB, CPY | adSOURCECODE |
| CWeb_Fmt | 768 | 709 | CWeb Source Code ⁴ | | W | adSOURCECODE |
| CartoCSS_Fmt | 769 | 710 | CartoCSS Source Code ⁴ | | MSS | adSOURCECODE |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|------------------------------|--------|----------|---|-------------------------|------------|--------------|
| Ceylon_Fmt | 770 | 711 | Ceylon Source Code ⁴ | text/x-ceylon | CEYLON | adSOURCECODE |
| Chapel_Fmt | 771 | 712 | Chapel Source Code ⁴ | | CHPL | adSOURCECODE |
| Clarion_Fmt | 772 | 713 | Clarion Source Code ⁴ | | CLW | adSOURCECODE |
| Clean_Fmt | 773 | 714 | Clean Source Code ⁴ | | DCL, ICL | adSOURCECODE |
| Component_Pascal_Fmt | 774 | 715 | Component Pascal Source Code ⁴ | text/x-component-pascal | CP | adSOURCECODE |
| Cool_Fmt | 775 | 716 | Cool Source Code ⁴ | | CL | adSOURCECODE |
| Coq_Fmt | 776 | 717 | Coq Source Code ⁴ | text/x-coq | V | adSOURCECODE |
| Creole_Fmt | 777 | 718 | Creole Source Code ⁴ | | CREOLE | adSOURCECODE |
| Crystal_Fmt | 778 | 719 | Crystal Source Code ⁴ | | CR | adSOURCECODE |
| Csound_Fmt | 779 | 720 | Csound Source Code ⁴ | | ORC | adSOURCECODE |
| Csound_Document_Fmt | 780 | 721 | Csound Document Source Code ⁴ | | CSD | adSOURCECODE |
| Cuda_Fmt | 781 | 722 | Cuda Source Code ⁴ | text/x-cuda | CU | adSOURCECODE |
| D_Fmt | 782 | 723 | D Source Code ⁴ | text/x-d | DCL, ICL | adSOURCECODE |
| DIGITAL_Command_Language_Fmt | 783 | 724 | DIGITAL Command Language Source Code ⁴ | | COM | adSOURCECODE |
| DTrace_Fmt | 784 | 725 | DTrace Source Code ⁴ | | D | adSOURCECODE |
| Dart_Fmt | 785 | 726 | Dart Source Code ⁴ | text/x-dart | DART | adSOURCECODE |
| E_Fmt | 786 | 727 | E Source Code ⁴ | | E | adSOURCECODE |
| ECL_Fmt | 787 | 728 | ECL Source Code ⁴ | application/x-ecl | ECL | adSOURCECODE |
| Elm_Fmt | 788 | 729 | Elm Source Code ⁴ | text/x-elm | ELM | adSOURCECODE |
| Emacs_Lisp_Fmt | 789 | 730 | Emacs Lisp Source Code ⁴ | text/x-emacs-lisp | EL | adSOURCECODE |
| EmberScript_Fmt | 790 | 731 | EmberScript Source Code ⁴ | | EM | adSOURCECODE |
| Fantom_Fmt | 791 | 732 | Fantom Source Code ⁴ | application/x-fantom | FAN | adSOURCECODE |
| Forth_Fmt | 792 | 733 | Forth Source Code ⁴ | text/x-forth | FOR, FORTH | adSOURCECODE |
| FreeMarker_Fmt | 793 | 734 | FreeMarker Source Code ⁴ | | FTL | adSOURCECODE |
| Frege_Fmt | 794 | 735 | Frege Source Code ⁴ | | FR | adSOURCECODE |
| G_code_Fmt | 795 | 736 | G-code Source Code ⁴ | | G | adSOURCECODE |
| GAMS_Fmt | 796 | 737 | GAMS Source Code ⁴ | | GMS | adSOURCECODE |
| GAP_Fmt | 797 | 738 | GAP Source Code ⁴ | | | adSOURCECODE |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|-------------------------|--------|----------|--|-----------------|-----------|--------------|
| GDScript_Fmt | 798 | 739 | GDScript Source Code ⁴ | | GD | adSOURCECODE |
| GLSL_Fmt | 799 | 740 | GLSL Source Code ⁴ | text/x-glslsrc | GLSL | adSOURCECODE |
| Game_Maker_Language_Fmt | 800 | 741 | Game Maker Language Source Code ⁴ | | GML | adSOURCECODE |
| Gnuplot_Fmt | 801 | 742 | Gnuplot Source Code ⁴ | text/x-gnuplot | GNU, GP | adSOURCECODE |
| Golo_Fmt | 802 | 743 | Golo Source Code ⁴ | | GOLO | adSOURCECODE |
| Gosu_Fmt | 803 | 744 | Gosu Source Code ⁴ | text/x-gosu | GS | adSOURCECODE |
| Gradle_Fmt | 804 | 745 | Gradle Source Code ⁴ | | GRADLE | adSOURCECODE |
| GraphQL_Fmt | 805 | 746 | GraphQL Source Code ⁴ | | GRAPHQL | adSOURCECODE |
| Graphviz_DOT_Fmt | 806 | 747 | Graphviz (DOT) Source Code ⁴ | | DOT | adSOURCECODE |
| HLSL_Fmt | 807 | 748 | HLSL Source Code ⁴ | | HLSL | adSOURCECODE |
| Hack_Fmt | 808 | 749 | Hack Source Code ⁴ | | | adSOURCECODE |
| Haml_Fmt | 809 | 750 | Haml Source Code ⁴ | text/x-haml | HAML | adSOURCECODE |
| Handlebars_Fmt | 810 | 751 | Handlebars Source Code ⁴ | | HBS | adSOURCECODE |
| Hy_Fmt | 811 | 752 | Hy Source Code ⁴ | text/x-hy | HY | adSOURCECODE |
| IDL_Fmt | 812 | 753 | IDL Source Code ⁴ | text/x-idl | PRO | adSOURCECODE |
| IGOR_Pro_Fmt | 813 | 754 | IGOR Pro Source Code ⁴ | text/ipf | IPF | adSOURCECODE |
| Idris_Fmt | 814 | 755 | Idris Source Code ⁴ | text/x-idris | IDR | adSOURCECODE |
| Inform_7_Fmt | 815 | 756 | Inform 7 Source Code ⁴ | | I7X | adSOURCECODE |
| Ioke_Fmt | 816 | 757 | Ioke Source Code ⁴ | text/x-iokesrc | IK | adSOURCECODE |
| Isabelle_Fmt | 817 | 758 | Isabelle Source Code ⁴ | text/x-isabelle | | adSOURCECODE |
| J_Fmt | 818 | 759 | J Source Code ⁴ | text/x-j | IJS | adSOURCECODE |
| JSONiq_Fmt | 819 | 760 | JSONiq Source Code ⁴ | | JQ | adSOURCECODE |
| JSX_Fmt | 820 | 761 | JSX Source Code ⁴ | | JSX | adSOURCECODE |
| Jasmin_Fmt | 821 | 762 | Jasmin Source Code ⁴ | | J | adSOURCECODE |
| Jolie_Fmt | 822 | 763 | Jolie Source Code ⁴ | | | adSOURCECODE |
| Julia_Fmt | 823 | 764 | Julia Source Code ⁴ | text/x-julia | JL | adSOURCECODE |
| KiCad_Layout_Fmt | 824 | 765 | KiCad Layout Source Code ⁴ | | | adSOURCECODE |
| KiCad_Schematic_Fmt | 825 | 766 | KiCad Schematic Source Code ⁴ | | SCH | adSOURCECODE |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|------------------|--------|----------|---------------------------------------|-------------------|------------|--------------|
| Kotlin_Fmt | 826 | 767 | Kotlin Source Code ⁴ | | KT | adSOURCECODE |
| LFE_Fmt | 827 | 768 | LFE Source Code ⁴ | text/x-kotlin | LFE | adSOURCECODE |
| LOLCODE_Fmt | 828 | 769 | LOLCODE Source Code ⁴ | | LOL | adSOURCECODE |
| Lasso_Fmt | 829 | 770 | Lasso Source Code ⁴ | text/x-lasso | LAS, LASSO | adSOURCECODE |
| Limbo_Fmt | 830 | 771 | Limbo Source Code ⁴ | text/limbo | | adSOURCECODE |
| LiveScript_Fmt | 831 | 772 | LiveScript Source Code ⁴ | text/x-livescript | LS | adSOURCECODE |
| M_Fmt | 832 | 773 | M Source Code ⁴ | | M | adSOURCECODE |
| MAXScript_Fmt | 833 | 774 | MAXScript Source Code ⁴ | | MS | adSOURCECODE |
| Markdown_Fmt | 834 | 775 | Markdown Source Code ⁴ | | MD | adSOURCECODE |
| Matlab_Fmt | 835 | 463 | Matlab Source Code ⁴ | text/x-matlab | M | adSOURCECODE |
| Max_Code_Fmt | 836 | 776 | Max Source Code ⁴ | | MXT | adSOURCECODE |
| Mercury_Fmt | 837 | 777 | Mercury Source Code ⁴ | | | adSOURCECODE |
| Modelica_Fmt | 838 | 778 | Modelica Source Code ⁴ | text/x-modelica | MO | adSOURCECODE |
| Modula_2_Fmt | 839 | 779 | Modula-2 Source Code ⁴ | text/x-modula2 | MOD | adSOURCECODE |
| Monkey_Fmt | 840 | 780 | Monkey Source Code ⁴ | text/x-monkey | MONKEY | adSOURCECODE |
| Moocode_Fmt | 841 | 781 | Moocode Source Code ⁴ | text/x-moocode | MOO | adSOURCECODE |
| NL_Fmt | 842 | 782 | NL Source Code ⁴ | | NL | adSOURCECODE |
| NSIS_Fmt | 843 | 783 | NSIS Source Code ⁴ | text/x-nsis | NSI | adSOURCECODE |
| NetLogo_Fmt | 844 | 784 | NetLogo Source Code ⁴ | | NLOGO | adSOURCECODE |
| NewLisp_Fmt | 845 | 785 | NewLisp Source Code ⁴ | text/x-newlisp | NL | adSOURCECODE |
| Nginx_Fmt | 846 | 786 | Nginx Source Code ⁴ | text/x-nginx-conf | VHOST | adSOURCECODE |
| Nix_Fmt | 847 | 787 | Nix Source Code ⁴ | text/x-nix | NIX | adSOURCECODE |
| Nu_Fmt | 848 | 788 | Nu Source Code ⁴ | | NU | adSOURCECODE |
| OCaml_Fmt | 849 | 789 | OCaml Source Code ⁴ | text/x-ocaml | | adSOURCECODE |
| OpenCL_Fmt | 850 | 790 | OpenCL Source Code ⁴ | | CL | adSOURCECODE |
| OpenEdge_ABL_Fmt | 851 | 791 | OpenEdge ABL Source Code ⁴ | text/x-openedge | | adSOURCECODE |
| OpenSCAD_Fmt | 852 | 792 | OpenSCAD Source Code ⁴ | | SCAD | adSOURCECODE |
| Ox_Fmt | 853 | 793 | Ox Source Code ⁴ | | OX | adSOURCECODE |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|---------------------|--------|----------|--|--------------------------|------------|--------------|
| Oxygene_Fmt | 854 | 794 | Oxygene Source Code ⁴ | | OXYGENE | adSOURCECODE |
| Oz_Fmt | 855 | 795 | Oz Source Code ⁴ | | OZ | adSOURCECODE |
| PAWN_Fmt | 856 | 796 | PAWN Source Code ⁴ | text/x-pawn | PWN | adSOURCECODE |
| PLpgSQL_Fmt | 857 | 797 | PLpgSQL Source Code ⁴ | text/x-plpgsql | PLSQL | adSOURCECODE |
| Pan_Fmt | 858 | 798 | Pan Source Code ⁴ | | PAN | adSOURCECODE |
| Parrot_Assembly_Fmt | 859 | 799 | Parrot Assembly Source Code ⁴ | | PASM | adSOURCECODE |
| PicoLisp_Fmt | 860 | 800 | PicoLisp Source Code ⁴ | | | adSOURCECODE |
| Pike_Fmt | 861 | 801 | Pike Source Code ⁴ | text/x-pike | PIKE | adSOURCECODE |
| Pony_Fmt | 862 | 802 | Pony Source Code ⁴ | | PONY | adSOURCECODE |
| Processing_Fmt | 863 | 803 | Processing Source Code ⁴ | | PDE | adSOURCECODE |
| PureBasic_Fmt | 864 | 804 | PureBasic Source Code ⁴ | | PB | adSOURCECODE |
| QMake_Fmt | 865 | 805 | QMake File ⁴ | | | adSOURCECODE |
| RAML_Fmt | 866 | 806 | RAML Source Code ⁴ | | RAML | adSOURCECODE |
| RDoc_Fmt | 867 | 807 | RDoc Source Code ⁴ | | RDOC | adSOURCECODE |
| REXX_Fmt | 868 | 808 | REXX Source Code ⁴ | text/x-rexx | REXX | adSOURCECODE |
| Racket_Fmt | 869 | 809 | Racket Source Code ⁴ | text/x-racket | | adSOURCECODE |
| Ragel_Fmt | 870 | 810 | Ragel Source Code ⁴ | | | adSOURCECODE |
| Rascal_Fmt | 871 | 811 | Rascal Source Code ⁴ | | RSC | adSOURCECODE |
| Rebol_Fmt | 872 | 812 | Rebol Source Code ⁴ | text/x-rebol | REB, REBOL | adSOURCECODE |
| Red_Fmt | 873 | 813 | Red Source Code ⁴ | text/x-red | RED | adSOURCECODE |
| RenPy_Fmt | 874 | 814 | Ren'Py Source Code ⁴ | | RPY | adSOURCECODE |
| RenderScript_Fmt | 875 | 815 | RenderScript Source Code ⁴ | | RS | adSOURCECODE |
| Ring_Fmt | 876 | 816 | Ring Source Code ⁴ | | RING | adSOURCECODE |
| RobotFramework_Fmt | 877 | 817 | RobotFramework Source Code ⁴ | text/x-robotframework | ROBOT | adSOURCECODE |
| SAS_Fmt | 878 | 818 | SAS Source Code ⁴ | | SAS | adSOURCECODE |
| SPARQL_Fmt | 879 | 819 | SPARQL format ⁴ | application/sparql-query | | adSOURCECODE |
| SQL_Fmt | 880 | 820 | SQL format ⁴ | text/x-sql | | adSOURCECODE |
| SQLPL_Fmt | 881 | 821 | SQLPL Source Code ⁴ | | | adSOURCECODE |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|-----------------------|--------|----------|---|---------------------------|-----------|-----------------|
| SaltStack_Fmt | 882 | 822 | SaltStack Source Code ⁴ | | SLS | adSOURCECODE |
| Scheme_Fmt | 883 | 823 | Scheme Source Code ⁴ | text/x-scheme | | adSOURCECODE |
| Scilab_Fmt | 884 | 824 | Scilab Source Code ⁴ | text/scilab | SCI | adSOURCECODE |
| Squirrel_Fmt | 885 | 825 | Squirrel Source Code ⁴ | | NUT | adSOURCECODE |
| Stan_Fmt | 886 | 826 | Stan Source Code ⁴ | | STAN | adSOURCECODE |
| Stata_Fmt | 887 | 827 | Stata Source Code ⁴ | | | adSOURCECODE |
| Stylus_Fmt | 888 | 828 | Stylus Source Code ⁴ | | STYL | adSOURCECODE |
| SuperCollider_Fmt | 889 | 829 | SuperCollider Source Code ⁴ | text/supercollider | SC | adSOURCECODE |
| SystemVerilog_Fmt | 890 | 830 | SystemVerilog Source Code ⁴ | text/x-systemverilog | SV | adSOURCECODE |
| TXL_Fmt | 891 | 831 | TXL Source Code ⁴ | | TXL | adSOURCECODE |
| Turing_Fmt | 892 | 832 | Turing Source Code ⁴ | | T | adSOURCECODE |
| Turtle_Fmt | 893 | 833 | Turtle Source Code ⁴ | text/turtle | TTL | adSOURCECODE |
| UrWeb_Fmt | 894 | 834 | UrWeb Source Code ⁴ | | UR, URS | adSOURCECODE |
| Vim_script_Fmt | 895 | 835 | Vim script File ⁴ | text/x-vim | VIM | adSOURCECODE |
| Visual_Basic_Fmt | 896 | 836 | Visual Basic Source Code ⁴ | text/x-vbasic | VB | adSOURCECODE |
| WebAssembly_Fmt | 897 | 837 | WebAssembly Source Code ⁴ | | WAT | adSOURCECODE |
| WebIDL_Fmt | 898 | 838 | WebIDL Source Code ⁴ | | WEBIDL | adSOURCECODE |
| X10_Fmt | 899 | 839 | X10 Source Code ⁴ | text/x-x10 | X10 | adSOURCECODE |
| XQuery_Fmt | 900 | 840 | XQuery Source Code ⁴ | text/xquery | XQM | adSOURCECODE |
| Xojo_Fmt | 901 | 841 | Xojo Source Code ⁴ | | | adSOURCECODE |
| Xtend_Fmt | 902 | 842 | Xtend Source Code ⁴ | text/x-xtend | XTEND | adSOURCECODE |
| YANG_Fmt | 903 | 843 | YANG Source Code ⁴ | | YANG | adSOURCECODE |
| Zephir_Fmt | 904 | 844 | Zephir Source Code ⁴ | | ZEP | adSOURCECODE |
| eC_Fmt | 905 | 845 | eC Source Code ⁴ | text/x-ecsrc | EC | adSOURCECODE |
| reStructuredText_Fmt | 906 | 846 | reStructuredText Source Code ⁴ | text/x-rst | | adSOURCECODE |
| xBase_Fmt | 907 | 847 | xBase Source Code ⁴ | | | adSOURCECODE |
| Windows_Installer_Fmt | 908 | 848 | MSI Windows Installer format | application/x-ole-storage | MSI | adENCAPSULATION |
| Autodesk_3ds_Max_Fmt | 909 | 849 | Autodesk 3ds Max format | | MAX | adCAD |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|-----------------------------------|--------|----------|--|---|-----------|-----------------|
| PhotoDraw_Mix_Fmt | 910 | 850 | PhotoDraw MIX image | image/vnd.mix | MIX | adRASTERIMAGE |
| Softimage_SCN_Fmt | 911 | 851 | Softimage Scene SCN format | | SCN | adCAD |
| Parasolid_XT_Fmt | 912 | 852 | Parasolid ascii XT format | | X_T | adCAD |
| Parasolid_XB_Fmt | 913 | 853 | Parasolid binary XB format | | X_B | adCAD |
| IGES_Fmt | 914 | 854 | Initial Graphics Exchange Specification format | model/iges | IGS | adCAD |
| ACE_Archive_Fmt | 915 | 855 | ACE archive format | application/x-ace-compressed | ACE | adENCAPSULATION |
| Grasshopper_GHX_Fmt | 916 | 856 | Grasshopper GHX format | | GHX | adCAD |
| MS_FrontPage_Macro_Fmt | 917 | 857 | Microsoft FrontPage macro file format | | FPM | adWORDPROCESSOR |
| MS_AtWork_Fax_Fmt | 918 | 858 | Microsoft AtWork Fax format | | AWD | adFAXFORMAT |
| MS_Image_Composer_Fmt | 919 | 859 | Microsoft Image Composer format | | MIC | adRASTERIMAGE |
| MS_Visual_InterDev_Fmt | 920 | 860 | Microsoft Visual InterDev web project items file | | WDM | adMISC |
| Macromedia_Flash_FLA_OLE_Fmt | 921 | 861 | Macromedia Flash FLA Project File OLE format | | FLA | adWORDPROCESSOR |
| Corel_Draw_X4_Fmt | 922 | 862 | CorelDRAW version X4 onwards | application/x-vnd.corel.zcf.draw.document+zip | CDRX | adVECTORGRAPHIC |
| Ogg_Daala_Fmt | 923 | 863 | Ogg Daala video format | video/daala | OGV | adMOVIE |
| Ogg_BBC_Dirac_Fmt | 924 | 864 | Ogg BBC Dirac video format | video/x-dirac | OGV | adMOVIE |
| PKCS_7_Fmt | 925 | 865 | PKCS #7 cryptographic format | application/pkcs7-signature | P7S | adWORDPROCESSOR |
| Time_Stamped_Data_Fmt | 926 | 866 | Time-stamped data format | application/timestamped-data | TSD | adENCAPSULATION |
| Sereal_Fmt | 927 | 867 | Sereal data serialization format | application/sereal | SRL | adMISC |
| Associated_Signature_Simple_Fmt | 928 | 868 | Associated Signature Container Simple format | application/vnd.etsi.asic-s+zip | ASICS | adENCAPSULATION |
| Associated_Signature_Extended_Fmt | 929 | 869 | Associated Signature Container Extended format | application/vnd.etsi.asic-e+zip | ASICE | adENCAPSULATION |
| iBooks_Fmt | 930 | 870 | Apple iBooks format | application/x-ibooks+zip | IBOOKS | adWORDPROCESSOR |
| PDF_Forms_Data_Fmt | 931 | 871 | PDF Forms Data Format | application/vnd.fdf | FDF | adWORDPROCESSOR |
| PDF_XML_Forms_Data_Fmt | 932 | 872 | PDF XML Forms Data Format | application/vnd.adobe.xfdf | XFDF | adWORDPROCESSOR |
| AxCrypt_Fmt | 933 | 873 | AxCrypt encrypted document | application/x-axcrypt | AXX | adENCAPSULATION |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|-----------------------------|--------|----------|--|------------------------------|------------|-----------------|
| Unix_Archive_Fmt | 934 | 874 | Unix Archive ar format | application/x-archive | AR | adENCAPSULATION |
| Berkeley_Btree_Database_Fmt | 935 | 875 | Berkeley DB btree database format | application/x-berkeley-db | DB | adDATABASE |
| Berkeley_Hash_Database_Fmt | 936 | 876 | Berkeley DB hash database format | application/x-berkeley-db | DB | adDATABASE |
| Berkeley_Log_Database_Fmt | 937 | 877 | Berkeley DB log database format | application/x-berkeley-db | | adDATABASE |
| Berkeley_Queue_Database_Fmt | 938 | 878 | Berkeley DB queue database format | application/x-berkeley-db | | adDATABASE |
| BitTorrent_Fmt | 939 | 879 | BitTorrent file format | application/x-bittorrent | TORRENT | adMISC |
| Chrome_Extension_Fmt | 940 | 880 | Google Chrome Extension format | application/x-chrome-package | CRX | adENCAPSULATION |
| Dalvik_Executable_Fmt | 941 | 881 | Dalvik Executable dex format | application/x-dex | DEX | adEXECUTABLE |
| Foxmail_Fmt | 942 | 882 | Foxmail email format | application/x-foxmail | BOX | adWORDPROCESSOR |
| GRIB_Fmt | 943 | 883 | General Regularly-distributed Information in Binary form GRIB format | application/x-grib | GRB, GRIB2 | adMISC |
| Zstandard_Fmt | 944 | 884 | Zstandard compression format | application/zstd | ZSTD | adENCAPSULATION |
| LZ4_Fmt | 945 | 885 | LZ4 compressed file | application/x-lz4 | LZ4 | adENCAPSULATION |
| MS_Money_Fmt | 946 | 886 | Microsoft Money format | application/x-msmoney | MNY | adSPREADSHEET |
| NetCDF_Fmt | 947 | 887 | Network Common Data Form NetCDF format | application/x-netcdf | NC | adMISC |
| SAS6_Data_Fmt | 948 | 888 | SAS 6 Data storage format | application/x-sas-data-v6 | SD2 | adDATABASE |
| SAS_Transport_Fmt | 949 | 889 | SAS Transport File XPORT format | application/x-sas-xport | XPT, XPORT | adDATABASE |
| Snappy_Framed_Fmt | 950 | 890 | Snappy Framed compression format | application/x-snappy-framed | SZ | adENCAPSULATION |
| Stata_Data_Fmt | 951 | 891 | Stata Data Format | application/x-stata-dta | DTA | adDATABASE |
| SPSS_SAV_Fmt | 952 | 892 | SPSS Statistics Data File Format | | SAV | adDATABASE |
| Zoo_Archive_Fmt | 953 | 893 | Zoo Compressed Archive Format | application/x-zoo | ZOO | adENCAPSULATION |
| CDX_Fmt | 954 | 894 | ChemDraw CDX format | chemical/x-cdx | CDX | adSCIENTIFIC |
| CDXML_Fmt | 955 | 895 | ChemDraw CDXML format | application/vnd.chemdraw+xml | CDXML | adSCIENTIFIC |
| BPG_Fmt | 956 | 896 | Better Portable Graphics BPG format | image/x-bpg | BPG | adRASTERIMAGE |
| Apple_Icon_Fmt | 957 | 897 | Apple Icon image format | image/icns | ICNS | adRASTERIMAGE |
| NITF_Fmt | 958 | 898 | National Imagery Transmission | image/nitf | NTF, NITF | adRASTERIMAGE |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|-------------------------------|--------|----------|---|-------------------------------|---------------|-----------------|
| | | | Format NITF image | | | |
| ERDAS_Imagine_Fmt | 959 | 899 | ERDAS Imagine image format | application/x-erdas-hfa | HFA, RRD, AUX | adRASTERIMAGE |
| MS_Office_Temporary_Owner_Fmt | 960 | 900 | Microsoft Office temporary owner file | application/x-ms-owner | | adMISC |
| EAC3_Audio_Fmt | 961 | 901 | Enhanced-AC3 (EAC3) Audio File format | audio/eac3 | AC3 | adSOUND |
| COFF_Relocatable_Fmt | 962 | 902 | Common Object File Format (COFF) relocatable object | application/x-object-file | O | adOBJECTMODULE |
| COFF_Executable_Fmt | 963 | 903 | Common Object File Format (COFF) executable | application/x-executable-file | | adEXECUTABLE |
| COFF_Dynamic_Lib_Fmt | 964 | 904 | Common Object File Format (COFF) dynamic library | application/x-library-file | | adLIBRARY |
| ELF_Core_Fmt | 965 | 905 | ELF Core file | application/x-coredump | | adMISC |
| Purify_Fmt | 966 | 906 | Rational Purify data file | | PFY | adMISC |
| Kryptel_Fmt | 967 | 907 | Kryptel encrypted file | | EDC | adENCAPSULATION |
| Windows_Core_Dump_Fmt | 968 | 908 | Windows heap or mini core dump file | application/x-dmp | DMP | adMISC |
| Qt_Prerendered_Font_Fmt | 969 | 909 | Qt Prerendered Font format | | QPF2 | adFONT |
| AIX_Relocatable_Fmt | 970 | 910 | AIX/RISC COFF relocatable object | application/x-object-file | | adOBJECTMODULE |
| AIX_Executable_Fmt | 971 | 911 | AIX/RISC COFF executable | application/x-executable-file | | adEXECUTABLE |
| AIX_Dynamic_Lib_Fmt | 972 | 912 | AIX/RISC COFF dynamic library | application/x-library-file | A | adLIBRARY |
| HPUX_Relocatable_Fmt | 973 | 913 | HPUX/PA-RISC COFF relocatable object | application/x-object-file | | adOBJECTMODULE |
| HPUX_Executable_Fmt | 974 | 914 | HPUX/PA-RISC COFF executable | application/x-executable-file | | adEXECUTABLE |
| HPUX_Dynamic_Lib_Fmt | 975 | 915 | HPUX/PA-RISC COFF dynamic library | application/x-library-file | SL | adLIBRARY |
| XML_EBCDIC_Fmt | 976 | 916 | EBCDIC-encoded XML file | application/xml | XML | adWORDPROCESSOR |
| MPEG_JVT_H264_Fmt | 977 | 917 | MPEG JVT-NAL sequence H264 video | video/h264 | 264 | adMOVIE |
| Material_Exchange_Fmt | 978 | 918 | Material Exchange Format audio-video container format | application/mxf | MXF | adMOVIE |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|-----------------------------|--------|----------|---|---|-----------|------------------|
| MS_Agent_Character_Fmt | 979 | 919 | Microsoft Agent Character file | | ACS | adMOVIE |
| Quicken_Fmt | 980 | 920 | Quicken data file | | QDF | adMISC |
| MS_Outlook_Address_Fmt | 981 | 921 | Microsoft Outlook address file | | WAB | adMISC |
| MS_Answer_Wizard_Fmt | 982 | 922 | Microsoft Answer Wizard file | | | adMISC |
| ADX_Fmt | 983 | 923 | ADX audio file | | ADX | adSOUND |
| System_Deployment_Image_Fmt | 984 | 924 | Microsoft System Deployment Image SDI format | | SDI | adMISC |
| Free_Lossless_Image_Fmt | 985 | 925 | Free Lossless Image Format (FLIF) | image/flif | FLIF | adRASTERIMAGE |
| DPX_Fmt | 986 | 926 | Digital Picture Exchange (DPX) image format | image/dpx | DPX | adRASTERIMAGE |
| Avro_Fmt | 987 | 927 | Apache Avro binary format | | AVRO | adMISC |
| InstallShield_Archive_Fmt | 988 | 928 | InstallShield archive (early versions) format | | EX_ | adENCAPSULATION |
| Mac_Executable_Fmt | 989 | 929 | Mac OS-X (Mach-O) executable format | | | adEXECUTABLE |
| GDSII_Fmt | 990 | 930 | GDSII data format | | GDS | adMISC |
| ActiveMime_Fmt | 991 | 931 | Microsoft ActiveMime (mso) documents | application/x-mso | MSO | adMISC |
| SmartCharts_Fmt | 992 | 932 | BizInt SmartCharts data format | | CHP, CHRR | adMISC |
| Webex_ARF_Fmt | 993 | 933 | Webex advanced network ARF recordings | | ARF | adMOVIE |
| Webex_WRF_Fmt | 994 | 934 | Webex local WRF recordings | | WRF | adMOVIE |
| PGP_NetShare_Fmt | 995 | 935 | Symantec PGP NetShare encrypted file | | | adENCAPSULATION |
| Ability_WP_OLE_Fmt | 996 | 936 | Ability Write later versions format | | AWW | adWORDPROCESSOR |
| Ability_SS_OLE_Fmt | 997 | 937 | Ability Spreadsheet later versions format | | AWS | adSPREADSHEET |
| InDesign_IDML_Fmt | 998 | 938 | Adobe InDesign IDML format | application/vnd.adobe.indesign-idml-package | IDML | adDESKTOPPUBLISH |
| Executable_JAR_Fmt | 999 | 939 | Executable Java Archive (jar) file | application/java-archive | JAR | adENCAPSULATION |
| IDOL_IDX_Fmt | 1000 | 940 | IDOL Server IDX file | | IDX | adENCAPSULATION |
| Android_Package_Kit_Fmt | 1001 | 941 | Android Package Kit (APK) format | application/vnd.android.package-archive | APK | adEXECUTABLE |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|-----------------------------|--------|----------|--|-------------------------------|-----------|-----------------|
| Android_Binary_XML_Fmt | 1002 | 942 | Android Binary XML (compressed by aapt) format | application/xml | XML | adWORDPROCESSOR |
| Java_WAR_Fmt | 1003 | 943 | Java WAR file format | | WAR | adENCAPSULATION |
| Java_EAR_Fmt | 1004 | 944 | Java EAR file format | | EAR | adENCAPSULATION |
| Atom_Syndication_Fmt | 1005 | 945 | Atom Syndication Format | application/atom+xml | ATOM | adWORDPROCESSOR |
| RSS_Fmt | 1006 | 946 | RSS syndication XML format | application/rss+xml | RSS | adWORDPROCESSOR |
| SMIL_Fmt | 1007 | 947 | Synchronized Multimedia Integration Language (SMIL) XML format | application/smil+xml | SMIL | adWORDPROCESSOR |
| XSLT_Fmt | 1008 | 948 | Extensible Stylesheet Language Transformations (XSLT) format | application/xslt+xml | XSL, XSLT | adWORDPROCESSOR |
| XML_Shareable_Playlist_Fmt | 1009 | 949 | XML Shareable Playlist Format (XSPF) | application/xspf+xml | XSPF | adWORDPROCESSOR |
| FictionBook_Fmt | 1010 | 950 | FictionBook e-book XML format | application/x-fictionbook+xml | FB2 | adWORDPROCESSOR |
| Adobe_Premiere_Project_Fmt | 1011 | 951 | Adobe Premiere project format | image/vnd.adobe.premiere | PPJ | adMISC |
| RDF_XML_Fmt | 1012 | 952 | RDF/XML format | application/rdf+xml | RDF | adWORDPROCESSOR |
| Really_Simple_Discovery_Fmt | 1013 | 953 | Really Simple Discovery (RSD) XML format | application/rsd+xml | RSD | adWORDPROCESSOR |
| SBML_Fmt | 1014 | 954 | Systems Biology Markup Language (SBML) XML format | application/sbml+xml | SBML | adWORDPROCESSOR |
| SRU_Fmt | 1015 | 955 | Search/Retrieve via URL (SRU) XML format | application/sru+xml | SRU | adWORDPROCESSOR |
| SSML_Fmt | 1016 | 956 | Speech Synthesis Markup Language (SSML) XML format | application/ssml+xml | SSML | adWORDPROCESSOR |
| PLS_Fmt | 1017 | 957 | Pronunciation Lexicon Specification (PLS) XML format | application/pls+xml | PLS | adWORDPROCESSOR |
| TEI_Fmt | 1018 | 958 | Text Encoding Initiative (TEI) XML format | application/tei+xml | TEI | adWORDPROCESSOR |
| METS_Fmt | 1019 | 959 | Metadata Encoding and Transmission Standard (METS) XML format | application/mets+xml | METS | adWORDPROCESSOR |
| MODS_Fmt | 1020 | 960 | Metadata Object Description Schema (MODS) XML format | application/mods+xml | MODS | adWORDPROCESSOR |
| Metalink_Fmt | 1021 | 961 | Metalink XML format | application/metalink4+xml | METALINK | adWORDPROCESSOR |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|------------------------------|--------|----------|--|---------------------------------|-----------|-----------------|
| Open_eBook_Fmt | 1022 | 962 | Open eBook (OEBPS) XML format | application/oebps-package+xml | OPF | adWORDPROCESSOR |
| SRGS_Fmt | 1023 | 963 | Speech Recognition Grammar Specification (SRGS) XML format | application/srgs+xml | SRGS | adWORDPROCESSOR |
| SPARQL_Results_Fmt | 1024 | 964 | SPARQL Query Results XML format | application/sparql-results+xml | SRX | adWORDPROCESSOR |
| Adobe_XML_Data_Package_Fmt | 1025 | 965 | Adobe XML Data Package format | application/vnd.adobe.xdp+xml | XDP | adWORDPROCESSOR |
| ESzigno_Fmt | 1026 | 966 | e-Szigno signed xml document | application/vnd.eszigno3+xml | ES3 | adWORDPROCESSOR |
| Mozilla_XUL_Fmt | 1027 | 967 | Mozilla XML User Interface Language (XUL) XML format | application/vnd.mozilla.xul+xml | XUL | adWORDPROCESSOR |
| SyncML_Fmt | 1028 | 968 | Synchronization Markup Language (SyncML) XML format | application/vnd.syncml+xml | XML | adWORDPROCESSOR |
| VoiceXML_Fmt | 1029 | 969 | VoiceXML (VXML) XML format | application/voicexml+xml | VXML | adWORDPROCESSOR |
| TI_Target_Configuration_Fmt | 1030 | 970 | Texas Instruments CCXML target configuration XML format | | CCXML | adWORDPROCESSOR |
| LZFSE_Fmt | 1031 | 971 | Lempel-Ziv Finite State Entropy (LZFSE) compression format | | LZFSE | adENCAPSULATION |
| Kindle_eBook_Fmt | 1032 | 972 | Amazon Kindle or Mobipocket eBook format | application/vnd.amazon.ebook | AZW, PRC | adWORDPROCESSOR |
| Oasis_Stream_Fmt | 1033 | 973 | Open Artwork System Interchange Standard (OASIS) format | | OAS | adMISC |
| Amazon_KFX_Fmt | 1034 | 974 | Amazon KFX eBook format | | KFX | adWORDPROCESSOR |
| KTX_Fmt | 1035 | 975 | KTX image format | image/ktx | KTX | adRASTERIMAGE |
| GMSH_Mesh_Fmt | 1036 | 976 | GMSH Mesh polygon format | model/mesh | MSH | adCAD |
| Collada_DAE_Fmt | 1037 | 977 | Collada Digital Asset Exchange (DAE) format | model/vnd.collada+xml | DAE | adCAD |
| YIN_Fmt | 1038 | 978 | YIN XML format | application/yin+xml | YIN | adWORDPROCESSOR |
| MPEG_Playlist_Fmt | 1039 | 979 | MPEG audio playlist format | audio/mpegurl | M3U | adSOUND |
| Windows_Audio_Playlist_Fmt | 1040 | 980 | Windows Audio playlist format | audio/x-ms-wax | WAX | adSOUND |
| DTS_Audio_Fmt | 1041 | 981 | DTS Coherent Acoustics audio format | audio/vnd.dts | DTS | adSOUND |
| Chemical_Markup_Language_Fmt | 1042 | 982 | Chemical Markup Language (CML) XML format | chemical/x-cml | CML | adWORDPROCESSOR |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|-------------------------------|--------|----------|--|--|-----------|-----------------|
| CrystalMaker_Fmt | 1043 | 983 | CrystalMaker chemical format | chemical/x-cmdf | CMDF | adSCIENTIFIC |
| VTK_XML_Fmt | 1044 | 984 | Visualization Toolkit VTK XML format | model/vnd.vtu | VTU | adVECTORGRAPHIC |
| IPFIX_Fmt | 1045 | 985 | IP Flow Information Export (IPFIX) format | application/ipfix | IPFIX | adMISC |
| Portable_Font_Resource_Fmt | 1046 | 986 | Portable Font Resource font format | application/font-tdpfr | PFR | adFONT |
| MARC_Fmt | 1047 | 987 | Machine-Readable Cataloging (MARC21) format | application/marc | MARC | adDATABASE |
| MARC_XML_Fmt | 1048 | 988 | Machine-Readable Cataloging (MARC) XML format | application/marcxml+xml | XML | adWORDPROCESSOR |
| XAR_Fmt | 1049 | 989 | Extensible Archive (XAR) format | | | adENCAPSULATION |
| Symbian_Installer_Fmt | 1050 | 990 | Symbian installer format | application/vnd.symbian.install | SIS | adENCAPSULATION |
| SO_Drawing_XML_Fmt | 1051 | 316 | OpenDocument format (OpenOffice 1/StarOffice 6.7) Drawing XML | application/vnd.sun.xml.draw | SXD | adVECTORGRAPHIC |
| SO_Text_Global_XML_Fmt | 1052 | 991 | OpenDocument format (OpenOffice 1/StarOffice 6.7) Writer Master document XML | application/vnd.sun.xml.writer.global | SXG | adWORDPROCESSOR |
| ODF_Chart_Fmt | 1053 | 992 | ODF Chart | application/vnd.oasis.opendocument.chart | ODC | adVECTORGRAPHIC |
| ODF_Database_Fmt | 1054 | 993 | ODF Database | application/vnd.sun.xml.base | ODB | adDATABASE |
| ODF_Image_Fmt | 1055 | 994 | ODF Image | application/vnd.oasis.opendocument.image | ODI | adRASTERIMAGE |
| ODF_Text_Master_Fmt | 1056 | 995 | ODF Text Master | application/vnd.oasis.opendocument.text-master | ODM | adWORDPROCESSOR |
| ODF_Text_Web_Fmt | 1057 | 996 | ODF Text Web | application/vnd.oasis.opendocument.text-web | OTH | adWORDPROCESSOR |
| ODF_Chart_Template_Fmt | 1058 | 997 | ODF Chart Template | application/vnd.oasis.opendocument.chart-template | OTC | adVECTORGRAPHIC |
| ODF_Formula_Template_Fmt | 1059 | 998 | ODF Formula Template | application/vnd.oasis.opendocument.formula-template | OTF | adWORDPROCESSOR |
| ODF_Drawing_Template_Fmt | 1060 | 316 | ODF Drawing/Graphics Template | application/vnd.oasis.opendocument.graphics-template | OTG | adVECTORGRAPHIC |
| ODF_Image_Template_Fmt | 1061 | 999 | ODF Image Template | application/vnd.oasis.opendocument.image-template | OTI | adRASTERIMAGE |
| ODF_Presentation_Template_Fmt | 1062 | 316 | ODF Presentation Template | application/vnd.oasis.opendocument.presentation-template | OTP | adPRESENTATION |
| ODF_Spreadsheet_Template_Fmt | 1063 | 315 | ODF Spreadsheet Template | application/vnd.oasis.opendocument.spreadsheet-template | OTS | adSPREADSHEET |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|------------------------------|--------|----------|---|---|-----------|-----------------|
| ODF_Text_Template_Fmt | 1064 | 314 | ODF Text Template | application/vnd.oasis.opendocument.text-template | OTT | adWORDPROCESSOR |
| ODF_Chart_XML_Fmt | 1065 | 1000 | ODF Chart flat XML format | application/vnd.oasis.opendocument.chart.xml | FODC | adVECTORGRAPHIC |
| ODF_Drawing_XML_Fmt | 1066 | 1001 | ODF Drawing/Graphics flat XML format | application/vnd.oasis.opendocument.formula.xml | FODG | adWORDPROCESSOR |
| ODF_Formula_XML_Fmt | 1067 | 1002 | ODF Formula flat XML format | application/vnd.oasis.opendocument.graphics.xml | FODF | adVECTORGRAPHIC |
| ODF_Image_XML_Fmt | 1068 | 1003 | ODF Image flat XML format | application/vnd.oasis.opendocument.image.xml | FODI | adRASTERIMAGE |
| ODF_Presentation_XML_Fmt | 1069 | 1004 | ODF Presentation flat XML format | application/vnd.oasis.opendocument.presentation.xml | FODP | adPRESENTATION |
| ODF_Spreadsheet_XML_Fmt | 1070 | 1005 | ODF Spreadsheet flat XML format | application/vnd.oasis.opendocument.spreadsheet.xml | FODS | adSPREADSHEET |
| ODF_Text_XML_Fmt | 1071 | 1006 | ODF Text flat XML format | application/vnd.oasis.opendocument.text.xml | FODT | adWORDPROCESSOR |
| ODF_Extension_Fmt | 1072 | 1007 | ODF Extension format | application/vnd.openofficeorg.extension | OXT | adMISC |
| StarView_Metafile_Fmt | 1073 | 1008 | OpenOffice StarView MetaFile format | image/x-svm | SVM | adRASTERIMAGE |
| BBeB_LRF_eBook_Fmt | 1074 | 1009 | Broad Band eBook (BBeB) in LRF format | | LRF | adWORDPROCESSOR |
| GPG_Trust_DB_Fmt | 1075 | 1010 | GPG trust database format | | GPG | adMISC |
| VICE_Emulator_Fmt | 1076 | 1011 | VICE (Versatile Commodore Emulator) format | | VSF | adMISC |
| Portable_Game_Notation_Fmt | 1077 | 1012 | Portable Game Notation chess format | application/vnd.chess-pgn | PGN | adWORDPROCESSOR |
| Doom_WAD_Fmt | 1078 | 1013 | Doom IWAD/PWAD format | application/x-doom | WAD | adMISC |
| Device_Tree_Blob_Fmt | 1079 | 1014 | Linux Device Tree Blob format | | DTB | adMISC |
| BDF_Font_Fmt | 1080 | 1015 | Glyph Bitmap Distribution Format | application/x-font-bdf | BDF | adFONT |
| PC_Screen_Font_Fmt | 1081 | 1016 | PC Screen Font format | application/x-font-psf | PSF | adFONT |
| JNLP_Fmt | 1082 | 1017 | Java Network Launching Protocol | application/x-java-jnlp-file | JNLP | adWORDPROCESSOR |
| XAML_Browser_Application_Fmt | 1083 | 1018 | XAML Browser Application (XBAP) format | application/x-ms-xbap | XBAP | adWORDPROCESSOR |
| MS_Binder_Fmt | 1084 | 1019 | Microsoft Office Binder format | application/x-msbinder | OBP | adENCAPSULATION |
| XAP_Fmt | 1085 | 1020 | Microsoft Silverlight application (XAP) format | application/x-silverlight-app | XAP | adENCAPSULATION |
| Stuftit_X_Fmt | 1086 | 1021 | Stuftit X (SITX) archive format | application/x-stuftitx | SITX | adENCAPSULATION |
| FIG_Fmt | 1087 | 1022 | Facility for Interactive Generation of figures (FIG) image format | application/x-xfig | FIG | adVECTORGRAPHIC |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|--------------------------|--------|----------|---|---|-----------|-----------------|
| XPIInstall_Fmt | 1088 | 1023 | XPIInstall Cross-Platform Installer Module (XPI) format | application/x-xpinstall | XPI | adENCAPSULATION |
| XDF_Fmt | 1089 | 1024 | Extensible Data Format (XDF) XML format | | XDF | adWORDPROCESSOR |
| MXML_Fmt | 1090 | 1025 | MXML UI markup language XML format | | MXML | adWORDPROCESSOR |
| MusicXML_Fmt | 1091 | 1026 | MusicXML format | application/vnd.recordare.musicxml | MXL | adENCAPSULATION |
| Finale_Fmt | 1092 | 1027 | Finale audio format | | MUS | adSOUND |
| Spotfire_DXP_Fmt | 1093 | 1028 | TIBCO Spotfire DXP data format | application/vnd.spotfire.dxp | DXP | adANALYTICS |
| MS_Office_Theme_2007_Fmt | 1094 | 1029 | Microsoft Office theme format | application/vnd.ms-officetheme | THMX | adMISC |
| Adobe_AIR_Installer_Fmt | 1095 | 1030 | Adobe AIR application installer package | application/vnd.adobe.air-application-installer-package+zip | AIR | adENCAPSULATION |
| Flex_Project_Fmt | 1096 | 1031 | Adobe Flash Flex project file format | application/vnd.adobe.fxp | FXP | adENCAPSULATION |
| FoxPro_Fmt | 1097 | 1032 | FoxPro compiled source format | | FXP | adLIBRARY |
| VST_Preset_Fmt | 1098 | 1033 | Virtual Studio Technology (VST) preset format | | FXP | adSOUND |
| Mischief_Image_Fmt | 1099 | 1034 | Mischief vector graphics image format | | ART | adVECTORGRAPHIC |
| FreeArc_Fmt | 1100 | 1035 | FreeArc archive format | application/x-freearc | ARC | adENCAPSULATION |
| Autodesk_3ds_Fmt | 1101 | 1036 | Autodesk 3ds format | application/x-3ds | 3DS | adCAD |
| Monkeys_Audio_Fmt | 1102 | 1037 | Monkey's Audio format | | APE | adSOUND |
| CALS_Fmt | 1103 | 1038 | CALS raster image format | | CAL | adRASTERIMAGE |
| Dr_Halo_PAL_Fmt | 1104 | 1039 | Dr Halo raster image PAL file format | | PAL | adRASTERIMAGE |
| DPG_Fmt | 1105 | 1040 | Nintendo DS DPG video format | | DPG | adMOVIE |
| JPEG_XR_Fmt | 1106 | 1041 | JPEG XR (extended range) image format | image/vnd.ms-photo | JXR, HDP | adRASTERIMAGE |
| TCR_eBook_Fmt | 1107 | 1042 | TCR (Text Compression for Reader) eBook format | | TCR | adWORDPROCESSOR |
| IHEX_Fmt | 1108 | 1043 | Intel Hex format | | IHEX | adENCAPSULATION |
| QCOW_Fmt | 1109 | 1044 | QEMU Copy On Write | | QCOW | adENCAPSULATION |
| VDI_Fmt | 1110 | 1045 | VirtualBox Disk Image | | VDI | adENCAPSULATION |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|---------------------------------|--------|----------|--|------------------------------------|-------------------------|------------------|
| OneNote_Alternate_Fmt | 1111 | 1046 | OneNote Alternative Packaging Format | | | adWORDPROCESSOR |
| RMS_Protected_Fmt | 1112 | 1047 | Rights Management Services (RMS)-protected format | | PFILE, PPDF, PJPG, PTXT | adENCAPSULATION |
| Portfolio_PDF_Fmt | 1113 | 1048 | Portfolio PDF File | application/pdf | PDF | adWORDPROCESSOR |
| Crystal_Reports_Fmt | 1114 | 1049 | SAP Crystal Reports format | application/x-rpt | RPT | adANALYTICS |
| Thumbs_db_Fmt | 1115 | 1050 | Microsoft Windows thumbs.db format | | DB | adENCAPSULATION |
| PagePlus_Fmt | 1116 | 1051 | Serif PagePlus format | | PPP | adDESKTOPPUBLISH |
| MS_Project_Exchange_Fmt | 1117 | 1052 | Microsoft Project Exchange format | | MPX | adSCHEDULE |
| MS_Management_Pack_MPX_Fmt | 1118 | 1053 | Microsoft Systems Center Operation Manager (SCOM) management pack MPX format | | MPX | adMISC |
| AutoCAD_VBA_Project_Fmt | 1119 | 1054 | AutoCAD VBA project format | | DVB | adMISC |
| PLY_ASCII_Fmt | 1120 | 1055 | Polygon File Format (PLY) ASCII format | | PLY | adCAD |
| PLY_Binary_Fmt | 1121 | 1056 | Polygon File Format (PLY) binary format | | PLY | adCAD |
| JavaView_JVX_Fmt | 1122 | 1057 | JavaView XML (JVX) format | | JVX | adCAD |
| X3D_Fmt | 1123 | 1058 | Extensible 3d Graphics (X3D) XML format | model/x3d+xml | X3D | adCAD |
| ZBrush_Project_Fmt | 1124 | 1059 | ZBrush ZProject (ZPR) format | | ZPR | adCAD |
| ZBrush_Tool_Fmt | 1125 | 1060 | ZBrush ZTool (ZTL) format | | ZTL | adCAD |
| Windows_Installer_Patch_Fmt | 1126 | 1061 | Microsoft Windows Installer Patch Package (MSP) format | | MSP | adENCAPSULATION |
| Windows_Installer_Transform_Fmt | 1127 | 1062 | Microsoft Windows Installer Transform (MST) format | | MST | adENCAPSULATION |
| Lotus_Approach_Fmt | 1128 | 1063 | Lotus Approach format | application/vnd.lotus-approach | APR, MPR | adDATABASE |
| Outlook_SendRcv_Settings_Fmt | 1129 | 1064 | Microsoft Outlook 2002 Send-Receive Settings | | SRS | adMISC |
| MS_Publisher_Scheme_Fmt | 1130 | 1065 | Microsoft Publisher colour scheme | | SCM | adMISC |
| SO_Chart_Fmt | 1131 | 1066 | Star Office 4,5 Chart | application/vnd.stardivision.chart | SDS | adVECTORGRAPHIC |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|----------------------------|--------|----------|---|-----------------------------------|-------------------------|------------------|
| SO_Database_Fmt | 1132 | 1067 | Star Office 4,5 Database | application/vnd.stardivision.base | SDB | adDATABASE |
| SO_Library_Fmt | 1133 | 1068 | Star Office 4,5 Library | | SBL | adLIBRARY |
| PageMaker_Document_Fmt | 1134 | 1069 | Adobe PageMaker document | application/pagemaker | PMD | adDESKTOPPUBLISH |
| MS_DTS_Fmt | 1135 | 1070 | Microsoft Data Transformation Services (DTS) package file | | DTS | adMISC |
| Cognos_PowerPlay_PPR_Fmt | 1136 | 1071 | Cognos PowerPlay up to version 7 (PPR) format | | PPR | adANALYTICS |
| Visual_Studio_SUO_Fmt | 1137 | 1072 | Microsoft Visual Studio solution user options (suo) file | | SUO | adMISC |
| MS_GraphEdit_Fmt | 1138 | 1073 | Microsoft GraphEdit File format | | GRF | adMISC |
| ArcGIS_Graph_Fmt | 1139 | 1074 | ArcGIS Graph format | | GRF | adGIS |
| SID_Audio_Fmt | 1140 | 1075 | SID Audio format | audio/prs.sid | SID | adSOUND |
| MrSID_Fmt | 1141 | 1076 | LizardTech MrSID image format | image/x-mrsid | SID | adRASTERIMAGE |
| Cardfile_Fmt | 1142 | 1077 | Microsoft Windows Cardfile address book format | application/x-mscardfile | CRD | adWORDPROCESSOR |
| MS_Word_Mac_4_Fmt | 1143 | 205 | Microsoft Word for Macintosh (version 4,5) | application/msword | DOC | adWORDPROCESSOR |
| WordPerfect_5_Fmt | 1144 | 80 | WordPerfect (version 5) | application/x-corel-wordperfect | WOP, DOC | adWORDPROCESSOR |
| WordPerfect_6_Fmt | 1145 | 178 | WordPerfect (version 6 and higher) | application/x-corel-wordperfect | WPD | adWORDPROCESSOR |
| WordPerfect_Graphics_1_Fmt | 1146 | 85 | WordPerfect Graphics (version 1) | application/vnd.wordperfect | WPG, QPG | AutoDetNoFormat |
| Organization_Chart_Fmt | 1147 | 1078 | OrgPlus Organization Chart | application/orgplus | OPX | adDATABASE |
| Lotus_Organizer_Fmt | 1148 | 1079 | Lotus Organizer documents | application/vnd.lotus-organizer | OR2, OR3, OR4, OR5, OR6 | adSCHEDULE |
| MS_DBML_Fmt | 1149 | 1080 | Microsoft Database Markup Language XML document | | DBML | adWORDPROCESSOR |
| XMind_Fmt | 1150 | 1081 | XMind document | application/xmind | XMIND | adPRESENTATION |
| MSI_Cerius_Fmt | 1151 | 1082 | MSI Cerius chemical formula document | chemical/x-cerius | MSI | adSCIENTIFIC |
| GenBank_Fmt | 1152 | 1083 | GenBank DNA character sequence document | chemical/x-genbank | GB | adSCIENTIFIC |
| GIS_World_File_Fmt | 1153 | 1084 | ESRI GIS World file | | BPW, GFW, JGW, J2W, | adGIS |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|---------------------------------|--------|----------|---|---------------------------------|------------------------------|-----------------|
| | | | | | PGW, SDW, TFW, WLD | |
| GIS_Projection_Metadata_Fmt | 1154 | 1085 | ESRI Projection Metadata (PRJ) file | | PRJ | adGIS |
| PowerWorld_Binary_Fmt | 1155 | 1086 | PowerWorld Binary (PWB) file | | PWB | adCAD |
| PowerWorld_Display_Fmt | 1156 | 1087 | PowerWorld Display (PWD) file | | PWD | adCAD |
| ArcXML_Fmt | 1157 | 1088 | ESRI ArcIMS project XML file (ArcXML) | | AXL | adGIS |
| GAMS_GDX_Fmt | 1158 | 1089 | General Algebraic Modeling System (GAMS) Data Exchange (GDX) format | | GDX | adSCIENTIFIC |
| ArcMap_MXD_Fmt | 1159 | 1090 | ArcMap Map Exchange Document project (MXD) | | MXD | adGIS |
| RRDtool_Fmt | 1160 | 1091 | RRDtool (Round Robin Database) data file | | RRD | adDATABASE |
| HWPX_Fmt | 1161 | 1092 | Hangul HWPX document | application/hwp+zip | HWPX | adWORDPROCESSOR |
| SolidWorks_2015_Fmt | 1162 | 1093 | SolidWorks (2015 onwards) file | | SLDPRT, SLDDRW, SLDASM | adCAD |
| MS_Photo_Editor_Fmt | 1163 | 1094 | Microsoft Photo Editor 'embedded GIF' file | application/vnd.ms-photo-editor | | adRASTERIMAGE |
| MS_Word_HTML_Fmt | 1164 | 1095 | Microsoft Word HTML format | | DOC, HTM | adWORDPROCESSOR |
| MS_Excel_HTML_Fmt | 1165 | 1096 | Microsoft Excel HTML format | | XLS, HTM | adWORDPROCESSOR |
| Portable_FloatMap_Fmt | 1166 | 1097 | Portable FloatMap (PFM) image | image/x-portable-floatmap | PFM | adRASTERIMAGE |
| RGBE_Fmt | 1167 | 1098 | Radiance RGBE (HDR) image | image/vnd.radiance | HDR, PIC, RGBE, XYZE | adRASTERIMAGE |
| APNG_Fmt | 1168 | 1099 | Animated Portable Network Graphics (Animated-PNG) | image/apng | APNG, PNG | adANIMATION |
| Enhanced_Compressed_Wavelet_Fmt | 1169 | 1100 | Enhanced Compressed Wavelet image | image/ecw | ECW | adRASTERIMAGE |
| Ensoniq_Waveset_Fmt | 1170 | 1101 | Ensoniq Waveset audio data file | | ECW | adSOUND |
| Corel_Photo_Paint_Fmt | 1171 | 1102 | Corel Photo Paint (version 7 and higher) | image/x-corelphotopaint | CPT | adRASTERIMAGE |
| OpenRaster_Fmt | 1172 | 1103 | OpenRaster image | image/openraster | ORA | adRASTERIMAGE |

| Format Name | Number | Category | Description | MIME Type | Extension | File Class |
|--------------------------|--------|----------|---|--|-----------|-----------------|
| Krita_Fmt | 1173 | 1104 | Krita image | application/x-krita | KRA | adRASTERIMAGE |
| Gerber_Fmt | 1174 | 1105 | Gerber image format | application/vnd.gerber | GBR | adVECTORGRAPHIC |
| PGML_Fmt | 1175 | 1106 | Precision Graphics Markup Language | | PGML | adVECTORGRAPHIC |
| Away3D_Fmt | 1176 | 1107 | Away3D scene file | | AWD | adCAD |
| CAD_3MF_Fmt | 1177 | 1108 | 3D Manufacturing Format document | application/vnd.ms-package.3dmanufacturing-3dmodel+xml | 3MF | adCAD |
| AMF_Fmt | 1178 | 1109 | Additive manufacturing file format (AMF) document | application/x-amf | AMF | adCAD |
| C3D_Fmt | 1179 | 1110 | Coordinate 3D (C3D) format | | C3D | adCAD |
| CAD_3DSystems_BFF_Fmt | 1180 | 1111 | 3D Sprint (3D Systems) SLA Build file | | BFF | adCAD |
| NRRD_Fmt | 1181 | 1112 | NRRD (nearly raw raster data) image format | | NRRD | adRASTERIMAGE |
| Cinema_4D_Fmt | 1182 | 1113 | Cinema 4D model | | C4D | adCAD |
| FBX_ASCII_Fmt | 1183 | 1114 | Kaydara FBX project (ASCII) | | FBX | adCAD |
| FBX_Binary_Fmt | 1184 | 1115 | Kaydara FBX project (binary) | | FBX | adCAD |
| Wavefront_OBJ_Fmt | 1185 | 1116 | Wavefront OBJ geometry definition file | | OBJ | adCAD |
| Wavefront_MTL_Fmt | 1186 | 1117 | Wavefront Material Template Library (MTL) | | MTL | adCAD |
| MS_Power_BI_Template_Fmt | 1187 | 1118 | Microsoft Power BI Desktop template format | | PBIT | adANALYTICS |

¹MHT, EML, and MBX files might return either format 2, 233, or 395, depending on the text in the file. In general, files that contain fields such as **To**, **From**, **Date**, or **Subject** are considered to be email messages; files that contain fields such as **content-type** and **mime-version** are considered to be MHT files; and files that do not contain any of those fields are considered to be text files.

²All CAT file extensions, for example CATDrawing, CATProduct, CATPart, and so on.

³This format is returned only if you enable source code identification. See [Source Code Identification, on page 107](#).

⁴This format is returned only if you enable extended source code identification. See [Source Code Identification, on page 107](#).

Appendix C: Character Sets

This section provides information on the handling of character sets in the KeyView suite of products, which includes KeyView Filter SDK, KeyView Export SDK, and KeyView Viewing SDK.

- [Multibyte and Bidirectional Support](#) 202
- [Coded Character Sets](#) 210

Multibyte and Bidirectional Support

The KeyView SDKs can process files that contain multibyte characters. A multibyte character encoding represents a single character with consecutive bytes. KeyView can also process text from files that contain bidirectional text. Bidirectional text contains both Latin-based text which is read from left to right, and text that is read from right to left (Hebrew and Arabic).

The following table indicates which character encodings are supported by KeyView for each format.

Multibyte and bidirectional support

| Format | Single-byte | Multibyte | Bidirectional |
|--|-------------|-----------|---------------|
| Archive | | | |
| 7-Zip (7Z) | n/a | n/a | n/a |
| AD1 Evidence file | n/a | n/a | n/a |
| ADJ | n/a | n/a | n/a |
| B1 | n/a | n/a | n/a |
| BinHex (HGX) | n/a | n/a | n/a |
| Bzip2 (BZ2) | n/a | n/a | n/a |
| EnCase – Expert Witness Compression Format (E01) | n/a | n/a | n/a |
| GZIP (GZ) | n/a | n/a | n/a |
| ISO (ISO) | n/a | n/a | n/a |
| Java Archive (JAR) | n/a | n/a | n/a |
| Legato EMailXtender Archive (EMX) | n/a | n/a | n/a |
| MacBinary (BIN) | n/a | n/a | n/a |
| Mac Disk Copy Disk Image (DMG) | n/a | n/a | n/a |
| Microsoft Backup File (BKF) | n/a | n/a | n/a |

Multibyte and bidirectional support, continued

| Format | Single-byte | Multibyte | Bidirectional |
|--|--------------------|------------------|----------------------|
| Microsoft Cabinet format (CAB) | n/a | n/a | n/a |
| Microsoft Compiled HTML Help (CHM) | n/a | n/a | n/a |
| Microsoft Compressed Folder (LZH) | n/a | n/a | n/a |
| PKZip (ZIP) | n/a | n/a | n/a |
| Microsoft Outlook DBX (DBX) | Y | Y | Y |
| Microsoft Outlook Offline Storage File (OST) | Y | Y | Y |
| RAR Archive (RAR) | n/a | n/a | n/a |
| Tape Archive (TAR) | n/a | n/a | n/a |
| UNIX Compress (Z) | n/a | n/a | n/a |
| UUEncoding (UUE) | n/a | n/a | n/a |
| Windows Scrap File (SHS) | n/a | n/a | n/a |
| WinZip (ZIP) | n/a | n/a | n/a |
| Binary | | | |
| Executable (EXE) | n/a | n/a | n/a |
| Link Library (DLL) | n/a | n/a | n/a |
| Computer-aided Design | | | |
| AutoCAD Drawing (DWG) | Y | Y | Y |
| AutoCAD Drawing Exchange (DXF) | Y | Y | Y |
| CATIA formats (CAT) | Y | N | N |
| Microsoft Visio (VSD) | Y | Y | Y |
| Database | | | |
| dBase Database | Y | N | N |
| Microsoft Access (MDB) | Y | Y | N |
| Microsoft Project (MPP) | Y | Y | N |
| Desktop Publishing | | | |
| Microsoft Publisher | N | Y | N |

Multibyte and bidirectional support, continued

| Format | Single-byte | Multibyte | Bidirectional |
|--|-------------|----------------|---------------|
| Display | | | |
| Adobe Portable Document Format (PDF) (basic reader) | Y | Y ¹ | Y |
| Adobe Portable Document Format (PDF) (graphic-based reader) | Y | Y ¹ | Y |
| Graphics | | | |
| Computer Graphics Metafile (CGM) | Y | N | N |
| Corel DRAW (CDR) | n/a | n/a | n/a |
| DCX Fax System (DCX) | Y | N | N |
| DICOM – Digital Imaging and Communications in Medicine (DCM) | n/a | n/a | n/a |
| Encapsulated PostScript (EPS) | Y | N | N |
| Enhanced Metafile (EMF) | Y | Y | N |
| Graphic Interchange Format (GIF) | n/a | n/a | n/a |
| JBIG2 | n/a | n/a | n/a |
| JPEG | n/a | n/a | n/a |
| JPEG 2000 | n/a | n/a | n/a |
| Lotus AMIDraw Graphics (SDW) | n/a | n/a | n/a |
| Lotus Pic (PIC) | n/a | n/a | n/a |
| Macintosh Raster (PICT/PCT) | n/a | n/a | n/a |
| MacPaint (PNTG) | n/a | n/a | n/a |
| Microsoft Office Drawing (MSO) | n/a | n/a | n/a |

¹Multibyte PDFs are supported, provided the PDF document is created by using either Character ID-keyed (CID) fonts, predefined CJK CMap files, or ToUnicode font encodings, and does not contain embedded fonts. See the Adobe website and the Adobe Acrobat documentation for more information. Any multibyte characters that are not supported are displayed using the replacement character. By default, the replacement character is a question mark (?).

To determine the type of font encodings that are used in a PDF, open the PDF in Adobe Acrobat, and select File > Document Info > Fonts. If the Encoding column lists Custom or Embedded encodings, you might encounter problems converting the PDF.

Multibyte and bidirectional support, continued

| Format | Single-byte | Multibyte | Bidirectional |
|---|--------------------|------------------|----------------------|
| Omni Graffle (GRAFFLE) | Y | N | N |
| PC PaintBrush (PCX) | n/a | n/a | n/a |
| Portable Network Graphics (PNG) | n/a | n/a | n/a |
| SGI RGB Image (RGB) | n/a | n/a | n/a |
| Sun Raster Image (RS) | n/a | n/a | n/a |
| Tagged Image File (TIFF) | Y | N | N |
| Truevision Targa (TGA) | n/a | n/a | n/a |
| Windows Animated Cursor (ANI) | n/a | n/a | n/a |
| Windows Bitmap (BMP) | n/a | n/a | n/a |
| Windows Icon Cursor (ICO) | n/a | n/a | n/a |
| Windows Metafile (WMF) | Y | Y | N |
| WordPerfect Graphics 1 (WPG) | Y | N | N |
| WordPerfect Graphics 2 (WPG) | Y | N | N |
| Mail | | | |
| Documentum EMCME Format | Y | Y | Y |
| Domino XML Language (DXL) | Y | Y | N |
| GroupWise FileSurf | Y | N | N |
| Legato Extender (ONM) | Y | Y | N |
| Lotus Notes database (NSF) | Y | Y | Y |
| Mailbox (MBX) | Y | Y | Y |
| Microsoft Entourage Database | Y | Y | Y |
| Microsoft Outlook (MSG) | Y | Y | Y |
| Microsoft Outlook Express (EML) | Y | Y | Y |
| Microsoft Outlook iCalendar | Y | Y | Y |
| Microsoft Outlook for Macintosh | Y | Y | Y |
| Microsoft Outlook Offline Storage File | Y | Y | Y |
| Microsoft Outlook Personal File Folders (PST) | Y | Y | Y |

Multibyte and bidirectional support, continued

| Format | Single-byte | Multibyte | Bidirectional |
|---|-------------------------|--|----------------------|
| Microsoft Outlook vCard Contact | | | |
| Text Mail (MIME) | Y | Y | Y |
| Transport Neutral Encapsulation Format | Y | Y | Y |
| Multimedia | | | |
| Advanced Systems Format (ASF) | n/a | n/a | n/a |
| Audio Interchange File Format (AIFF) | n/a | n/a | n/a |
| Microsoft Wave Sound (WAV) | n/a | n/a | n/a |
| MIDI (MID) | n/a | n/a | n/a |
| MPEG 1 Audio Layer 3 (MP3) | n/a | n/a | n/a |
| MPEG 1 Video (MPG) | n/a | n/a | n/a |
| MPEG 2 Audio (MPEGA) | n/a | n/a | n/a |
| MPEG 4 Audio (MP4) | n/a | n/a | n/a |
| NeXT/Sun Audio (AU) | n/a | n/a | n/a |
| QuickTime Movie (QT/MOV) | n/a | n/a | n/a |
| Windows Video (AVI) | n/a | n/a | n/a |
| Presentations | | | |
| Apple iWork Keynote (GZ) | Y | Y | N |
| Applix Presents (AG) | character set 1252 only | N | N |
| Corel Presentations (SHW) | character set 1252 only | N | N |
| Extensible Forms Description Language (XFD) | Y | Y | N |
| Lotus Freelance Graphics 2 (PRE) | character set 850 only | N | N |
| Lotus Freelance Graphics (PRZ) | Y | Japanese, Simple Chinese, Traditional Chinese, Thai only | N |
| Macromedia Flash (SWF) | Y | Y | N |

Multibyte and bidirectional support, continued

| Format | Single-byte | Multibyte | Bidirectional |
|---|-------------------------|--|----------------------|
| Microsoft OneNote | Y | Y | N |
| Microsoft PowerPoint PC (PPT) | character set 1252 only | Traditional Chinese only | N |
| Microsoft PowerPoint Windows (PPT) | Y | Japanese, Simple Chinese, Traditional Chinese, Korean only | Hebrew only |
| Microsoft PowerPoint Macintosh (PPT) | Y | N | N |
| Microsoft PowerPoint Windows XML 2007 and 2010 (PPTX) | Y | Y | Y |
| OASIS Open Document (ODP) | Y | Y | N |
| OpenOffice Impress (ODP) | Y | Y | N |
| StarOffice Impress (ODP) | Y | Y | N |
| Spreadsheets | | | |
| Apple iWork Numbers (GZ) | Y | Y | N |
| Applix Spreadsheets (AS) | character set 1252 only | N | N |
| Comma Separated Values (CSV) | character set 1252 only | N | N |
| Corel Quattro Pro (QPW/WB3) | Y | N | N |
| Data Interchange Format (DIF) | Y | Y | Y ¹ |
| Lotus 1-2-3 (123) | Y | Y | Y |
| Lotus 1-2-3 (WK4) | Y | Y | N |
| Lotus 123 Charts (123) | Y | Y | N |
| Microsoft Excel Charts (XLS) | Y | Y | N |
| Microsoft Excel Macintosh (XLS) | Y | N | N |
| Microsoft Excel Windows (XLS) | Y | Y | Y ² |
| Microsoft Excel Windows XML 2007 (XLSX) | Y | Y | N |
| Microsoft Office Excel Binary Format (XLSB) | Y | Y | N |
| Microsoft Works Spreadsheet | Y | N | N |

Multibyte and bidirectional support, continued

| Format | Single-byte | Multibyte | Bidirectional |
|--------------------------------------|------------------------------|------------------|----------------------|
| (S30/S40) | | | |
| OASIS Open Document (ODS) | Y | Y | N |
| OpenOffice Calc (ODS) | Y | Y | N |
| StarOffice Calc (ODS) | Y | Y | N |
| Text and Markup | | | |
| ANSI (TXT) | Y | Y | Y ² |
| ASCII (TXT) | Y | Y | Y ² |
| HTML (HTM) | Y | Y | Y ^{2, 2} |
| Microsoft Excel Windows XML 2003 | Y | Y | Y |
| Microsoft Word for Windows XML 2003 | Y | Y | Y |
| Microsoft Visio XML 2003 | Y | Y | Y |
| Rich Text Format (RTF) | Y | Y | Y ³ |
| Unicode HTML | Y | Y | Y ^{2, 3} |
| Unicode Text (TXT) | Y | Y | Y ² |
| XHTML | Y | Y | Y ³ |
| XML | Y | Y | Y |
| Word Processing | | | |
| Adobe Maker Interchange Format (MIF) | character set 1252 only | N | N |
| Apple iChat Log (ICHAT) | Y | Y | N |
| Apple iWork Pages (GZ) | Y | Y | N |
| Applix Words (AW) | character set 1252 only | N | N |
| DisplayWrite (IP) | character set 500, 1026 only | N | N |
| Folio Flat File (FFF) | character set 1252 only | N | N |
| Founder Chinese E-paper Basic (CEB) | Y | Y | N |

Multibyte and bidirectional support, continued

| Format | Single-byte | Multibyte | Bidirectional |
|--|----------------------------------|---|--------------------------|
| Fujitsu Oasys (OA2) | Y | Y | N |
| Hangul (HWP) | Y | Y | N |
| Health level7 (HL7) | Y | Y | Y |
| IBM DCA/RTF (DC) | character sets 500, 1026 only | N | N |
| JustSystems Ichitaro (JTD) | Y | Y | N |
| Lotus AMI Pro (SAM) | Y | Simple Chinese, Traditional Chinese, Japanese, Thai only | Y |
| Lotus AMI Professional Write Plus (AMI) | Y | Simple Chinese, Traditional Chinese, Japanese, Thai only | N |
| Lotus Word Pro (LWP) | Y | Y | Y ³ |
| Lotus SmartMaster (MWP) | Y | Y | N |
| Microsoft Word PC (DOC) | character set 1252 only | N | N |
| Microsoft Word Windows V1-2 (DOC) | Y | N | N |
| Microsoft Word Windows V6, 7, 8, 95 (DOC) | Y | Y | Hebrew only ³ |
| Microsoft Word Windows V97 through 2003 (DOC) | Y | Y | Y ³ |
| Microsoft Word Windows XML 2007 and 2010 (DOCX) | Y | Y | Y ³ |
| Microsoft Word Macintosh (DOC) | Y | N | Y ³ |
| Microsoft Works (WPS) | Y | Japanese only | N |
| Microsoft Write (WRI) | Y | Japanese only | N |
| OASIS Open Document (ODT) | Y | Y | N |
| Omni Outliner (OO3) | Y | Y | N |
| OpenOffice Writer (ODT) | Y | Y | N |
| Open Publication Structure eBook (EPUB) | Y | Y | Y |
| StarOffice Writer (ODT) | Y | Y | N |

Multibyte and bidirectional support, continued

| Format | Single-byte | Multibyte | Bidirectional |
|--------------------------------|----------------------------|------------------------------|---------------|
| Skype Log (DBB) | Y | Y (null-terminated charsets) | N |
| WordPad (RTF) | Y | Y | Y |
| WordPerfect Linux (WPS) | Y | N | N |
| WordPerfect Macintosh (WPS) | Y | N | N |
| WordPerfect Windows (WO) | Y | N | N |
| XML Paper Specification (XPS) | Y | Y | N |
| XYWrite Windows (XY4) | character set 1252 only | N | N |
| Yahoo! Instant Messenger (DAT) | Y | Y (null-terminated charsets) | N |

¹The text direction in the output file might not be correct.

²In Export SDK, a bidirectional right-to-left (RTL) tag is extracted from this format and included in the direction element (`<dir=RTL>`) of the output.

Coded Character Sets

This section lists which character set you can use to specify the target character set. The coded character sets are enumerated in `kvtypes.h` and defined in the Export class.

Code Character Sets

| Coded Character Set | Description | Can be set as target charset? |
|---------------------|---|-------------------------------|
| KVCS_UNKNOWN | Unknown character set | N |
| KVCS_SJIS | Japanese (uses multibyte encoding), cp932 | Y |
| KVCS_GB | Simplified Chinese (China, Singapore, Malaysia) cp936 | Y |
| KVCS_BIG5 | Traditional Chinese (Taiwan, Hong Kong, Macaw) cp950 | Y |
| KVCS_KSC | Korean, cp949 | Y |
| KVCS_1250 | Windows Latin 2 (Central Europe) | Y |
| KVCS_1251 | Windows Cyrillic (Slavic) | Y |

Code Character Sets, continued

| Coded Character Set | Description | Can be set as target charset? |
|----------------------------|--|--------------------------------------|
| KVCS_1252 | Windows Latin 1 (ANSI) | Y |
| KVCS_1253 | Windows Greek | Y |
| KVCS_1254 | Windows Latin 5 (Turkish) | Y |
| KVCS_1255 | Windows Hebrew | Y |
| KVCS_1256 | Windows Arabic | Y |
| KVCS_1257 | Windows Baltic Rim | Y |
| KVCS_1258 | Windows Vietnamese | Y |
| KVCS_8859_1 | ISO 8859-1 Latin 1 (Western Europe, Latin America) | Y |
| KVCS_8859_2 | ISO 8859-2 Latin 2 (Central Eastern Europe) | Y |
| KVCS_8859_3 | ISO 8859-3 Latin 3 (S.E. Europe) | Y |
| KVCS_8859_4 | ISO 8859-4 Latin 4 (Scandinavia/Baltic) | Y |
| KVCS_8859_5 | ISO 8859-5 Latin/Cyrillic | Y |
| KVCS_8859_6 | ISO 8859-6 Latin/Arabic | Y |
| KVCS_8859_7 | ISO 8859-7 Latin/Greek | Y |
| KVCS_8859_8 | ISO 8859-8 Latin/Hebrew | Y |
| KVCS_8859_9 | ISO 8859-9 Latin/Turkish | Y |
| KVCS_8859_14 | ISO 8859-14 | Y |
| KVCS_8859_15 | ISO 8859-15 | Y |
| KVCS_437 | DOS Latin US | Y |
| KVCS_737 | DOS Greek | Y |
| KVCS_775 | DOS Baltic Rim | Y |
| KVCS_850 | DOS Latin 1 | Y |
| KVCS_851 | DOS Greek | Y |
| KVCS_852 | DOS Latin 2 | Y |
| KVCS_855 | DOS Cyrillic | Y |

Code Character Sets, continued

| Coded Character Set | Description | Can be set as target charset? |
|----------------------------|---|--------------------------------------|
| KVCS_857 | DOS Turkish | Y |
| KVCS_860 | DOS Portuguese | Y |
| KVCS_861 | DOS Icelandic | Y |
| KVCS_862 | DOS Hebrew | Y |
| KVCS_863 | DOS Canadian French | Y |
| KVCS_864 | DOS Arabic | Y |
| KVCS_865 | DOS Nordic | Y |
| KVCS_866 | DOS Cyrillic Russian | Y |
| KVCS_869 | DOS Greek 2 | Y |
| KVCS_874 | Thai | Y |
| KVCS_PDFMACDOC | PDF MAC DOC | N |
| KVCS_PDFWINDOC | PDF WIN DOC | N |
| KVCS_STDENC | Adobe Standard Encoding | N |
| KVCS_PDFDOC | Adobe standard PDF character set | N |
| KVCS_037 | EBCDIC code page 037 | Y |
| KVCS_1026 | EBCDIC code page 1026 | Y |
| KVCS_500 | EBCDIC code page 500 | Y |
| KVCS_875 | EBCDIC code page 875 | Y |
| KVCS_LMBCS | Lotus multibyte character set Group 1 and Group 2 | N |
| KVCS_UNICODE | Unicode, UCS-2 | N |
| KVCS_UTF16 | 16-bit Unicode transformation format | N |
| KVCS_UTF8 | 8-bit Unicode transformation format | Y |
| KVCS_UTF7 | 7-bit Unicode transformation format | Y |
| KVCS_2022_JP | ISO 2022-JP, Japanese mail and news safe encoding (JIS-7) | N |

Code Character Sets, continued

| Coded Character Set | Description | Can be set as target charset? |
|----------------------------|---|--------------------------------------|
| KVCS_2022_CN | ISO 2022-CN, Chinese mail and news safe encoding | N |
| KVCS_2022_KR | ISO 2022-KR, Korean mail and news safe encoding | N |
| KVCS_WP6X | Word Perfect 6.x and higher character mapping | N |
| KVCS_10000 | Western European (Macintosh) | Y |
| KVCS_KSC5601 | Unified Hangul | Y |
| KVCS_GB2312 | Simplified Chinese (China, Singapore, Hong Kong) | Y |
| KVCS_GB12345 | Traditional Chinese (China) - analogue of GB2312 | Y |
| KVCS_CNS11643 | Traditional Chinese - Taiwan. Supplement to Big5 | Y |
| KVCS_JIS0201 | Japanese - contains ASCII character set (JIS-Roman) | N |
| KVCS_JIS0212 | Japanese. Supplement to JIS0208. | Y |
| KVCS_EUC_JP | Japanese Extended UNIX Code | Y |
| KVCS_EUC_GB | Simplified Chinese Extended UNIX Code | Y |
| KVCS_EUC_BIG5 | Traditional Chinese Extended UNIX Code | N |
| KVCS_EUC_KSC | Korean Extended UNIX Code | N |
| KVCS_424 | EBCDIC Hebrew | N |
| KVCS_856 | PC Hebrew (old) | N |
| KVCS_1006 | IBM AIX Pakistan (Urdu) | N |
| KVCS_KOI8R | Cyrillic (Russian) | Y |
| KVCS_PDF_JAPAN1 | Adobe-Japan1-2 character collection | N |
| KVCS_PDF_KOREA1 | Adobe-Korea1-0 character collection | N |
| KVCS_PDF_GB1 | Adobe-GB1-3 character collection | N |
| KVCS_PDF_ | Adobe-CNS1-2 character collection | N |

Code Character Sets, continued

| Coded Character Set | Description | Can be set as target charset? |
|----------------------------|--|--------------------------------------|
| CNS1 | | |
| KVCS_2022_JP_8 | ISO 2022-JP, Japanese mail and news safe encoding (JIS8) | N |
| KVCS_720 | Arabic DOS-720 | Y |
| KVCS_VISCII | Vietnamese VISCII | Y |
| KVCS_8859_10 | ISO 8859-10 (Latin 6 Nordic) | Y ¹ |
| KVCS_8859_13 | ISO 8859-13 (Latin 7 Baltic) | Y 1 |
| KVCS_57002 | ISCII Devanagari (x-iscii-de) | Y 1 |
| KVCS_57003 | ISCII Bengali (x-iscii-be) | Y 1 |
| KVCS_57004 | ISCII Tamil (x-iscii-ta) | Y1 |
| KVCS_57005 | ISCII Telugu (x-iscii-te) | Y1 |
| KVCS_57006 | ISCII Assamese (x-iscii-as) | Y1 |
| KVCS_57007 | ISCII Oriya (x-iscii-or) | Y1 |
| KVCS_57008 | ISCII Kannada (x-iscii-ka) | Y1 |
| KVCS_57009 | ISCII Malayalam (x-iscii-ma) | Y1 |
| KVCS_57010 | ISCII Gujarathi (x-iscii-gu) | Y1 |
| KVCS_57011 | ISCII Panjabi (x-iscii-pa) | Y 1 |
| KVCS_GB18030b2 | Reserved for internal use | n/a |
| KVCS_GB18030 | GB18030 (Chinese 4-byte character set) | Y |
| KVCS_8859_11 | ISO 8859-11 (Thai) | Y |
| KVCS_8859_16 | ISO 8859-16 (Latin-10 South-Eastern Europe) | Y |
| KVCS_ARABICMAC | Arabic Mac (x-mac-arabic) | Y |
| KVCS_KOI8U | Cyrillic (KOI8U Ukrainian) | Y |
| KVCS_HZGB2312 | The 7-bit representation of GB 2312 / RFC 1842 | n/a |

¹The character set cannot be forced as output in Export SDK and Viewing SDK because the character

set is not supported by the major browsers.

Appendix D: Extract and Format Lotus Notes Subfiles

This section describes how to create XML templates to alter the appearance of extracted Lotus mail note subfiles so that they maintain the look and feel of the original notes.

- [Overview](#)216
- [Customize XML Templates](#)216
- [Template Elements and Attributes](#)218
- [Date and Time Formats](#)223

Overview

KeyView uses the NSF reader, `nsfsr`, to extract Lotus database files, and places Lotus mail notes in subfiles. The NSF reader uses a set of default XML templates to extract the notes and apply formatting, thereby approximating the look and feel of the original notes.

In some cases, you might need to customize the XML templates, for instance if your notes contain custom data. In such cases, you can modify the existing XML templates or create your own.

During extraction, the NSF reader loads all XML files in the `NSFtemplates` directory and its subdirectories (except for the `NSFtemplates\images` directory, which is reserved for images). During initialization, the KeyView XML parser verifies the XML templates. If the templates contain any invalid XML, elements, or attributes, initialization fails and errors are recorded in the `nsfsr.log` file.

Customize XML Templates

XML templates are enabled by default. In most cases, the default templates should be sufficient; however, you can customize them or create your own as required.

To customize XML templates for Lotus note extraction

1. Modify the template files in the following directory.

`install\OS\bin\NSFtemplates`

The `main.xml` file must exist in the `NSFtemplates` directory. It is the top-level template file that extracts all subfiles, usually by calling other templates.

2. Make sure that any modifications or additional XML files conform to the supported elements and attributes described in [Template Elements and Attributes, on page 218](#).
3. Extract the Lotus database file.

Use Demo Templates

For testing purposes, you can extract notes by using a set of demo templates, which are provided to demonstrate the proper usage of all the XML elements and attributes, because the default templates do not use all the XML elements.

The demo templates are available at:

install\OS\bin\NSFtemplates

To use the demo XML templates

1. In the `formats.ini` file, set the following parameter.

```
[nsfsr]
UseDemoTemplate=1
```

2. In the `main.xml` file, uncomment the following section.

```
<ifini name="UseDemoTemplate" text="1">
  <call file="demo.xml"/>
  <quit/>
</ifini>
```

Use Old Templates

For testing purposes, you can extract notes by using legacy templates, which produce MHTML output. You can generate similar output by disabling the XML templates, but using the old templates enables you to see the XML code and compare it to the standard and demo templates.

To use the old XML templates

1. In the `formats.ini` file, set the following parameter.

```
[nsfsr]
UseOldTemplate=1
```

2. In the `main.xml` file, uncomment the following section.

```
<ifini name="UseOldTemplate" text="1">
  <call file="default_old.xml">
  <quit>
</ifini>
```

Disable XML Templates

For testing purposes, you can disable XML templates; KeyView extracts the notes in MHTML format. You can compare the MHTML output directly by the NSF reader with the MHTML output indirectly by the NSF reader through the XML templates.

To disable XML templates

1. In the `formats.ini` file, set the following parameter.

```
[nsfsr]  
ExtractByTemplate=0
```

Template Elements and Attributes

This section lists the valid XML elements and attributes that you can use when creating or modifying templates. See the demo templates for examples.

Conditional Elements

The following table lists the valid conditional elements.

Conditional elements

| Element | Description |
|--|---|
| <keyview> | The KeyView XML template container ("root") element |
| <if*> | <p>If the condition from the comparison is true, process the XML. Conditions can be nested up to 25 levels deep.</p> <p>Attributes</p> <ul style="list-style-type: none">• <code>name</code>. (Required) The name of the main item to compare to <code>item</code> or <code>text</code>.• <code>item</code>. (Required if no <code>text</code>) The name of the item to compare to the item specified by <code>name</code>.• <code>text</code>. (Required if no <code>item</code>) The text to compare to the item specified by <code>name</code>. |
| <ifex>, <ifnx> | <p>If <code>name</code> item exists and has a <code>text</code> value or not.</p> <p>The Notes item might have a value that cannot be converted to text, such as an image.</p> |
| <ifeq>, <ifne>, <iflt>, <ifle>, <ifgt>, <ifge> | <p>Respectively, if <code>text</code> ==, !=, <, >, <=, >, >=.</p> <p>Text comparison uses a case-insensitive string compare.</p> |
| <iftdeq>, <iftdne>, <iftdlt>, <iftdle>, <iftdgt>, <iftdge> | <p>Respectively, if time/date ==, !=, <, >, <=, >, >=.</p> <p>Time/date comparison converts dates to text in local time using the Notes default, <code>TZFMT_NEVER</code>, because Notes also sometimes converts fields to text internally. For example:</p> <p><code>text="06/30/2005 02:52:04 PM"</code></p> |

Conditional elements, continued

| Element | Description |
|--------------------|--|
| <iftzeq>, <iftzne> | Respectively, if the time zone equals or does not equal the comparison text, for example CDT, EST, and so on. |
| <ifini> | If the value of the INI option specified in name equals the text value. |
| <else> | If the condition from the last <if> or <switch> was false, process XML. |
| <switch> | <p>If a name value exists, process XML.</p> <p>Attributes</p> <ul style="list-style-type: none"> name. (Required) The name of the main item to compare in <case> subelements. |
| <case> | <p>If the comparison condition is true, process XML, then stop processing the rest of <switch>.</p> <p>Attributes</p> <ul style="list-style-type: none"> text. (Required) The text to compare to the name item of <switch>. |
| <default> | If all <case> conditions were false, process XML. This element must be the last element in <switch>, after all the <case> elements. Any <case> elements after the <default> element are ignored. |
| <for> | <p>If a name value exists, process XML. Process for each part of the name item.</p> <p>Attributes</p> <ul style="list-style-type: none"> name. (Required) The name of the main item. max. (Optional) The maximum index to process. By default, all are processed. |
| <index> | Output <for> loop index (1-based). <index> is only valid within a <for> element. |

Control Elements

The following table lists the valid control elements.

Control Elements

| Element | Description |
|---------|---|
| <call> | <p>Call another XML template. You can nest templates up to 10 levels deep.</p> <p>Attributes</p> |

Control Elements, continued

| Element | Description |
|---------------------------|--|
| | <ul style="list-style-type: none">• <code>file</code>. (Required) The template file name. This name must be unique. |
| <code><log></code> | <p>Log message to the NSF log file.</p> <p>Attributes</p> <ul style="list-style-type: none">• <code>text</code>. (Required) The text to log.• <code>type</code>. (Optional) The type of log message. The following values are valid:<ul style="list-style-type: none">◦ ERROR◦ WARN◦ INFO◦ DIAG (the default option)◦ DEBUG◦ DUMP |
| <code><quit></code> | <p>Stop processing the template. Exits without error.</p> <p>Attributes</p> <ul style="list-style-type: none">• <code>text</code>. (Optional) The text to log.• <code>type</code>. (Optional) The type of log message. See <log>, above. |
| <code><stop></code> | <p>Stop processing the template. Exits with an ERROR log message.</p> <p>Attributes</p> <ul style="list-style-type: none">• <code>text</code>. (Required) The text to log. |

Data Elements

The following table lists the valid data elements.

Data elements

| Element | Description |
|---------------------------|--|
| <code><text></code> | <p>Output text.</p> <p>Attributes</p> <ul style="list-style-type: none">• <code>name</code>. (Required if there is no parent) The name of the item to output. |
| <code><rich></code> | <p>Output rich text (MHTML). Images are output in the next part or parts of the MHTML, after the first <code><HTML></code> part.</p> |

Data elements, continued

| Element | Description |
|----------|---|
| | Attributes <ul style="list-style-type: none"> name. (Required if there is no parent) The name of the item to output. |
| <body> | Output the message body in rich text (MHTML). As with <rich> , on the previous page, images are output in the next part or parts of the MHTML. |
| <form> | Output the message form (usually \$Body field) in rich text (MHTML). Attributes <ul style="list-style-type: none"> name. (Required if there is no parent) The name of the item to output. |
| <addr> | Output an address. Attributes <ul style="list-style-type: none"> name. (Required if there is no parent) The name of the item to output. type. (Optional) The type of address to output. Set this attribute to CN (Common Name), which is the only supported type. |
| <name> | Output the name of the last name item, or in other words the current main item. The item must exist. |
| <format> | Set the default format for <date> and <date_kv>. This element does not set the <text> format. See Date and Time Formats, on page 223 for a list of all Notes and KeyView date and time formats and integer values. Attributes <ul style="list-style-type: none"> format. (Optional. Omit to reset to defaults) The Notes and KeyView date and time format. You can set the following formats: <ul style="list-style-type: none"> TD=int. The Time Date format (TDFMT_*) TS=int. The Time Show format (TSFMT_*) TT=int. The Time Time format (TTFMT_*) TZ=int. The Time Zone format (TZFMT_*) KV=int. The KeyView date and time format <p>where int is an integer value that corresponds to the desired format.</p> <p>Separate multiple formats with commas. For example:</p> <p>format="TD=0,TS=2,TT=1,TZ=1,KV=55"</p> |
| <date> | Output a Notes date. Attributes <ul style="list-style-type: none"> name. (Required if there is no parent) The name of the item to output. |

Data elements, continued

| Element | Description |
|-------------|---|
| | <ul style="list-style-type: none"> format. (Optional) See <format>, on the previous page. You can set the following values: <ul style="list-style-type: none"> TD TS TT TZ |
| <date_kv> | <p>Output a KeyView date.</p> <p>Attributes</p> <ul style="list-style-type: none"> name. (Required if there is no parent) The name of the item to output. format. (Optional) See <format>, on the previous page. You can set the following values: <ul style="list-style-type: none"> TZ KV |
| <time> | <p>Output a time range, for example 1 hour, 30 minutes.</p> <p>Attributes</p> <ul style="list-style-type: none"> name. (Required if there is no parent) The item name of the start date or time. item. (Required) The item name of the end date or time. |
| <zone> | <p>Output a Notes time zone mnemonic, for example MST.</p> <p>Attributes</p> <ul style="list-style-type: none"> name. (Required if there is no parent) The name of date item to output. |
| <zone_utc> | <p>Output a time zone as UTC, for example (UTC-06:00).</p> |
| <logo> | <p>Output the mail header logo.</p> <p>The image link is included in the output; the actual image is output to a different part of the MHTML subfile.</p> |
| <image> | <p>Output an image.</p> <p>The image link is included in the output; the actual image is output to the MHTML next part, as with <rich>, on page 220 and <body>, on the previous page.</p> |
| <image_uri> | <p>Output an image URI, in quotation marks. The actual image is output to a different part of the MHTML subfile.</p> <p>Attributes</p> |

Data elements, continued

| Element | Description |
|---------|--|
| | <ul style="list-style-type: none">• <code>link</code>. (Required if there is no <code>file</code>) The image link, such as a form or title name. For example:<ul style="list-style-type: none">• <code>link="StdNotesLtr0"</code>• <code>file</code>. (Required if there is no <code>link</code>) The name of the image file. The file must exist in the <code>.././templates/images</code> directory. For example:<ul style="list-style-type: none">• <code>file="boxcheck.gif"</code> |

Date and Time Formats

This section lists the supported Notes and KeyView date and time formats for use with `<format>`, `<date>`, and `<date_kv>`.

Lotus Notes Date and Time Formats

This section lists supported Lotus Notes date and time formats, and the integer values that specify each one.

Lotus Notes date and time formats

| Format | Integer Value | Description |
|-----------------|---------------|---|
| TDFMT_FULL | 0 | (The Notes default) Year, month, and day |
| TDFMT_CPARTIAL | 1 | Month and day, year if not this year |
| TDFMT_PARTIAL | 2 | Month and day |
| TDFMT_DPARTIAL | 3 | Year and month |
| TDFMT_FULL4 | 4 | Four-digit year, month, and day |
| TDFMT_CPARTIAL4 | 5 | Month and day, four-digit year if not this year |
| TDFMT_DPARTIAL4 | 6 | Four-digit year and month |
| TTFMT_FULL | 0 | (Notes default) Hour, minute, and second |
| TTFMT_PARTIAL | 1 | Hour and minute |
| TTFMT_HOUR | 2 | Hour |

Lotus Notes date and time formats, continued

| Format | Integer Value | Description |
|-----------------|---------------|---|
| TZFMT_NEVER | 0 | (Notes default) All time zones are converted to the current time zone |
| TZFMT_SOMETIMES | 1 | Show only when outside the current time zone |
| TZFMT_ALWAYS | 2 | Show for all time zones |
| TSFMT_DATE | 0 | Date |
| TSFMT_TIME | 1 | Time |
| TSFMT_DATETIME | 2 | (The Notes default) Date and time |
| TSFMT_CDATETIME | 4 | Date and time, or time today or time yesterday |

KeyView Date and Time Formats

This section lists KeyView date and time formats. The KeyView formats use the following syntax:

| | |
|---------|---|
| Month | Month = full month name Mon = abbreviated month name m = month (number) mm = two-digit month (leading 0) |
| Weekday | Weekday = full weekday name Wday = abbreviated weekday name |
| Year | yy = two-digit year yyyy = four-digit year |
| >Day | d = day (number) dd = two-digit day (leading 0) |
| Time | h = 12-hour H = 24-hour m = minutes s = seconds P = AM/PM p = am/pm |

Separators _ = space
 c = comma
 s = slash
 a = dash
 o = dot

KeyView date and time formats

| Format | Output | Integer Value |
|--|------------|---------------|
| 12-Hour and 24-Hour Time Formats | | |
| KVDTF_P | P | 1 |
| KVDTF_P_hmm | P h:mm | 2 |
| KVDTF_hmm_P | h:mm P | 3 |
| KVDTF_P_hhmm | P hh:mm | 4 |
| KVDTF_hhmm_P | hh:mm P | 5 |
| KVDTF_P_hmmss | P h:mm:ss | 6 |
| KVDTF_hmmss_P | h:mm:ss P | 7 |
| KVDTF_P_hhmmss | P hh:mm:ss | 8 |
| KVDTF_hhmmss_P | hh:mm:ss P | 9 |
| KVDTF_Hmm | H:mm | 10 |
| KVDTF_HHmm | HH:mm | 11 |
| KVDTF_mmss | mm:ss | 12 |
| KVDTF_Hmmss | H:mm:ss | 13 |
| KVDTF_HHmmss | HH:mm:ss | 14 |
| Numerical Date Formats with Slashes | | |
| KVDTF_mmsdd | mm/dd | 15 |
| KVDTF_msdsyy | m/d/yy | 16 |
| KVDTF_mmsddsyy | mm/dd/yy | 17 |
| KVDTF_mmsddsyyyy | mm/dd/yyyy | 18 |
| KVDTF_ddsmm | dd/mm | 19 |

KeyView date and time formats, continued

| Format | Output | Integer Value |
|-------------------------|---------------------|---------------|
| KVDTF_ddsmsyy | dd/mm/yy | 20 |
| KVDTF_ddsmsyy_Hmm | dd/mm/yy H:mm | 21 |
| KVDTF_ddsmm_P_hmm | dd/mm P h:mm | 22 |
| KVDTF_ddsmm_hmm_P | dd/mm h:mm P | 23 |
| KVDTF_ddsmm_P_hhmm | dd/mm P hh:mm | 24 |
| KVDTF_ddsmm_hhmm_P | dd/mm hh:mm P | 25 |
| KVDTF_ddsmsyy_P_hmm | dd/mm/yy P h:mm | 26 |
| KVDTF_ddsmsyy_hmm_P | dd/mm/yy h:mm P | 27 |
| KVDTF_ddsmsyy_P_hmmss | dd/mm/yy P h:mm:ss | 28 |
| KVDTF_ddsmsyy_hmmss_P | dd/mm/yy h:mm:ss P | 29 |
| KVDTF_ddsmsyy_P_hhmmss | dd/mm/yy P hh:mm:ss | 30 |
| KVDTF_ddsmsyy_hhmmss_P | dd/mm/yy hh:mm:ss P | 31 |
| KVDTF_yysmmsdd_P_hhmmss | yy/mm/dd P hh:mm:ss | 32 |
| KVDTF_yysmmsdd_hhmmss_P | yy/mm/dd hh:mm:ss P | 33 |
| KVDTF_msdsyy_Hmm | m/d/yy H:mm | 34 |
| KVDTF_mmsddsyy_Hmm | mm/dd/yy H:mm | 35 |
| KVDTF_msdsyy_P_hmm | m/d/yy P h:mm | 36 |
| KVDTF_msdsyy_hmm_P | m/d/yy h:mm P | 37 |
| KVDTF_mmsddsyy_hmm_P | mm/dd/yy h:mm P | 38 |
| KVDTF_mmsdd_P_hhmm | mm/dd P hh:mm | 39 |
| KVDTF_mmsdd_hhmm_P | mm/dd hh:mm P | 40 |
| KVDTF_mmsddsyy_P_hhmmss | mm/dd/yy P hh:mm:ss | 41 |
| KVDTF_mmsddsyy_hhmmss_P | mm/dd/yy hh:mm:ss P | 42 |
| KVDTF_msd | m/d | 43 |
| KVDTF_yysm | yy/m | 44 |
| KVDTF_yysmm | yy/mm | 45 |

KeyView date and time formats, continued

| Format | Output | Integer Value |
|--|---------------------|---------------|
| KVDTF_ysmsd | yy/m/d | 46 |
| KVDTF_ysmmsdd | yy/mm/dd | 47 |
| KVDTF_yysmmsdd | yyyy/mm/dd | 48 |
| Numerical Date Formats with Dashes | | |
| KVDTF_ddammayy | dd-mm-yy | 49 |
| KVDTF_mmadd | mm-dd | 50 |
| KVDTF_mmayy | mm-yy | 51 |
| KVDTF_yyammadd | yy-mm-dd | 52 |
| KVDTF_yyyymmadd | yyyy-mm-dd | 53 |
| KVDTF_yyyymmaddaHHmmss | yyyy-mm-dd-HH:mm:ss | 54 |
| Numerical Date Formats with Dots | | |
| KVDTF_yyomod | yy.m.d | 55 |
| KVDTF_yyommodd | yy.mm.dd | 56 |
| KVDTF_mod | m.d | 57 |
| KVDTF_mmodd | mm.dd | 58 |
| Numerical and String Date Formats with Dashes, Commas, and Spaces | | |
| KVDTF_ddaMon | dd-Mon | 59 |
| KVDTF_daMonayy | d-Mon-yy | 60 |
| KVDTF_ddaMonayy | dd-Mon-yy | 61 |
| KVDTF_ddaMonayyyy | dd-Mon-yyyy | 62 |
| KVDTF_Mon | Mon | 63 |
| KVDTF_Monayy | Mon-yy | 64 |
| KVDTF_Monayyyy | Mon-yyyy | 65 |
| KVDTF_Monaddayy | Mon-dd-yy | 66 |
| KVDTF_yyammadd_P_hhmmss | yy-mm-dd P hh:mm:ss | 67 |
| KVDTF_mmadd_P_hhmm | mm-dd P hh:mm | 68 |

KeyView date and time formats, continued

| Format | Output | Integer Value |
|------------------------------|------------------------|---------------|
| KVDTF_Mon_yy | Mon yy | 69 |
| KVDTF_Monc_yy | Mon, yy | 70 |
| KVDTF_Month | Month | 71 |
| KVDTF_Monthayy | Month-yy | 72 |
| KVDTF_Month_yy | Month yy | 73 |
| KVDTF_Monthc_yy | Month, yy | 74 |
| KVDTF_Monthayyyy | Month-yyyy | 75 |
| KVDTF_Month_yyyy | Month yyyy | 76 |
| KVDTF_Monthc_yyyy | Month, yyyy | 77 |
| KVDTF_Mon_dc_yyyy | Mon d, yyyy | 78 |
| KVDTF_d_Monc_yyyy | d Mon, yyyy | 79 |
| KVDTF_yyyy_Mon_d | yyyy Mon d | 80 |
| KVDTF_Month_dc_yyyy | Month d, yyyy | 81 |
| KVDTF_d_Monthc_yyyy | d Month, yyyy | 82 |
| KVDTF_yyyy_Month_d | yyyy Month d | 83 |
| Weekday Date Formats | | |
| KVDTF_wday | wday | 84 |
| KVDTF_Weekday | Weekday | 85 |
| KVDTF_wdayc_Mon_dc_yyyy | wday, Mon d, yyyy | 86 |
| KVDTF_Weekdayc_Month_dc_yyyy | Weekday, Month d, yyyy | 87 |
| KVDTF_Weekdayc_d_Monthc_yyyy | Weekday, d Month, yyyy | 88 |

Appendix E: Export Tokens

This section contains an alphabetized list of the Export tokens.

Tokens are special strings inserted into the `KVHTMLTemplateEx` structure, `HtmlTemplateInfo` class, and template files. They are placeholders for markup that appears in the HTML output. For example, the `$CHARSET` token marks the place in the HTML output where the name of the source document's character set is inserted. It would be used in the tag `<charset=$CHARSET>`.

Word documents are split into blocks by heading level. By default, each section of text between Heading Level 1 headings will be a single block.

See the template files for examples of how to use tokens.

Export Tokens

| Token | Description |
|----------------------------|---|
| <code>\$ANCHOR</code> | Inserts an anchor for a heading level (h2-h6) for the current block. |
| <code>\$BASE</code> | Inserts the base URL for the HTML file. Use in the <code><base href=xx></code> tag. |
| <code>\$CHARSET</code> | Inserts the character set of the source document, if that information is ascertainable. Supported Formats, on page 121 lists the file formats for which character set information can be determined. |
| <code>\$CONTENT</code> | Inserts the content of the metadata field specified by the <code>\$NAME</code> token. This token is used in conjunction with the <code>\$SUMMARY</code> , <code>\$USERSUMMARY</code> , and <code>\$NAME</code> tokens to insert source document metadata into the HTML output. An example of this token's use is: <code>pszUserSummary=<meta name="\$NAME" content="\$CONTENT"></code> Supported Formats, on page 121 lists file formats that support metadata. |
| <code>\$ENDNOTE</code> | Inserts endnotes from the current block at this point in the output stream. Currently implemented for Microsoft Word documents only. |
| <code>\$ENDNOTEALL</code> | Inserts all endnotes at this point in the output stream. Currently implemented for Microsoft Word documents only. |
| <code>\$FOOTER</code> | Inserts the footer from the current block at this point in the output stream. |
| <code>\$FOOTNOTE</code> | Inserts footnotes from the current block at this point in the output stream. Currently implemented for Microsoft Word documents only. |
| <code>\$FOOTNOTEALL</code> | Inserts all footnotes at this point in the output stream. Currently implemented for Microsoft Word documents only. |
| <code>\$HEADER</code> | Inserts the header from the current block at this point in the output stream. |

Export Tokens, continued

| Token | Description |
|--------------------|--|
| \$MAINURL | Inserts the URL to the file containing the start of the generated HTML, that is, the main output stream. |
| \$NAME | <p>Inserts the name of a metadata field. This token is used in conjunction with the \$SUMMARY, below, \$USERSUMMARY, on the next page, and \$CONTENT, on the previous page tokens to insert source document metadata into the HTML output. An example of this token's use is:</p> <pre>pszUserSummary=<meta name="\$NAME" content="\$CONTENT"></pre> <p>The section Supported Formats, on page 121 lists file formats that support metadata.</p> |
| \$NEXT | Inserts the anchor to the next block. If this is the last block, a link to the first block is inserted. |
| \$PREV | Inserts the anchor to the previous block. If the current block is the first block, a link to the last block is inserted. |
| \$STYLESHEET | Inserts the path to the style sheet. Only available in KVHTMLOptionsEx. |
| \$SUMMARY | <p>Inserts the data from standard metadata fields using the markup provided in the pszUserSummary member of the structure KVHTMLTemplateEx. Standard fields are enumerated from 0 to 33 in KVSUMType in kvtypes.h. See the tokens \$USERSUMMARY, on the next page, \$NAME, above, and \$CONTENT, on the previous page.</p> <p>The section Supported Formats, on page 121 lists file formats that support metadata.</p> |
| \$SUMMARYNN | <p>Inserts the data from a <i>specified</i> metadata field. <i>NN</i> is a number from 0 through 33 enumerated in the KVSUMType structure in kvtypes.h. An example of this token's use is:</p> <pre>pszMainTop=<head> <title> \$SUMMARY01 </title> </head> <body></pre> <p>The section Supported Formats, on page 121 lists file formats that support metadata.</p> |
| \$SPLITBLOCKNUMBER | Inserts the page number for each block generated as a result of bHardPageMakesNewBlock or lcbBlockSize. |
| \$TOC | Inserts the table of contents at this point in the current output stream. This token is typically embedded in pszMainTop. |
| \$TOCB | Inserts the table of contents at this point for the current block. |
| \$TOCBE | Inserts the beginning entry for the table of contents at this point in the current output stream. |

Export Tokens, continued

| Token | Description |
|---------------|---|
| \$TOCE | Inserts a table of contents entry at this point in the current output stream. |
| \$TOCTE | Inserts a text entry without HTML markup at this point in the current output stream. |
| \$TOCPE | Inserts a partial table of contents entry at this point in the current output stream. HTML tags are removed; however, character entities are retained. This enables angle brackets to appear in the table of contents entries (for example, <text>). Without this token, <text> would be interpreted as a non-valid HTML tag and would be ignored by the browser. |
| \$TOPANCHOR | Inserts the anchor for the top heading level (h1) for the current block. |
| \$USERCB | Triggers the callback function <code>UserCB()</code> and identifies the callback used in the function. |
| \$USERSUMMARY | <p>Inserts the data from every valid non-standard metadata field using the markup provided in the <code>pszUserSummary</code> member of the <code>KVHTMLTemplateEx</code> structure. Non-standard metadata are any fields not listed from 0 to 33 in <code>KVSumType</code>, such as user-defined fields (for example, custom property fields in Word documents), or fields that are unique to a particular file type (for example, "Artist" or "Genre" fields in MP3 files). See the tokens \$SUMMARY, on the previous page, \$NAME, on the previous page, and \$CONTENT, on page 229.</p> <p>The section Supported Formats, on page 121 lists file formats that support metadata.</p> |
| \$XANCHOR | <p>Inserts the anchor to an extra file into the HTML output. An example of this token's use is:</p> <pre><frame src="\$XANCHOR" name="Left" scrolling="auto" target="right"></pre> <p>The contents of the extra file is defined by <code>pszXFile</code>, and the block generated by this token is defined by <code>pszXStartBlock</code> and <code>pszXEndBlock</code>.</p> |

Appendix F: File Format Detection

This section describes how file formats are detected in the KeyView Export SDK.

| | |
|--|-----|
| • Introduction | 232 |
| • Extract Format Information | 232 |
| • Determine Format Support | 232 |
| • Translate Format Information | 234 |
| • Determine a Document Reader | 236 |
| • Category Values in formats_e.ini | 236 |

Introduction

The KeyView format detection module (`kwad`) detects a file's format, and reports the information to the API, which in turn reports the information to the developer's application. If the detected format is supported by the KeyView SDK, the detection module also loads the appropriate structured access layer and document reader for further processing.

For a list of supported formats, see [Supported Formats, on page 121](#).

Extract Format Information

You can extract format information from a document by using the `getAutoDetectInfo()` method. This method extracts the major format, file class, version, and document attributes, and populates the `AutoDetectInfo` class. It returns the same format information as the `fpGetStreamInfo()` function, but as a string not an integer. The format information that can be extracted is listed in the `adinfo.h` header file.

For information on how to translate the extracted format information, see [Translate Format Information, on page 234](#).

Determine Format Support

After the file format is extracted, the detection module uses the `formats_e.ini` file to determine whether the format is supported by KeyView, and the appropriate structured access layer and reader to load.

The `formats_e.ini` file is in the directory `install\OS\bin`, where `install` is the path name of the Export installation directory and `OS` is the name of the operating system. It contains the following information:

- Coded format information. To translate this information, see [Translate Format Information, on page 234](#).

- The reader associated with each format. See [Determine a Document Reader, on page 236](#).
- Configuration parameters for out-of-process conversions.
- Locale settings for internal use.

Below are some entries from the `formats_e.ini` file:

```
123=mw
152=xyw
178=wp6
189=mw6
2=af
200=pdf
205=mb
210=htm
251=htm
```

NOTE: The `formats_e.ini` file applies to all formats except graphics. Detection of graphics formats is handled by an internal module named KeyView Picture Interchange Format (KPIF).

Refine Detection of Text Files

During text detection, KeyView analyzes the first 1 kB and last 1 kB of data in a document; if less than 10% of that data consists of non-ASCII characters, KeyView detects the document as a text file.

However, depending on the type of documents you are working with, the default settings might not provide the desired level of accuracy. Configuration flags allow you to change the amount of data to read at the end of a file, the percentage of non-ASCII characters permitted in a text file, and whether to use or ignore the file extension to determine the document format.

Change the Amount of File Data to Read

During file detection, KeyView reads characters from the beginning and end of a file—by default, it reads the first and last 1,024 bytes of data. Large text files might contain many irrelevant characters at the end of a file, so KeyView might not accurately detect the file format. You can set a configuration flag to increase the amount of data to read from the end of a file during detection.

To change the amount of data to read during detection

- In the `formats_e.ini` file, set the following flag in the `detection_flags` section:

```
[detection_flags]
non_ascii_chars_end_block_size=kB
```

where *kB* is the number of kilobytes to read from the end of the file, from 0 to 10. The default value is 1.

NOTE: The file size must be greater than the value specified in the flag. If the flag value is greater than the file size, KeyView does not use the flag.

Change the Percentage of Allowed Non-ASCII Characters

By default, if less than 10% of the analyzed data in a document consists of non-ASCII characters, it is detected as a text file. Depending on the type of files you are working with, changing the default percentage might increase detection accuracy.

To change the percentage of non-ASCII characters allowed in text files

- In the `formats_e.ini` file, set the following flag in the `detection_flags` section:

```
[detection_flags]
non_ascii_chars_in_text=N
```

where *N* is the percentage of non-ASCII characters to allow in text files. Files that contain a lower percentage of non-ASCII characters than *N* are detected as text files. The default value is 10.

Use the File Extension for Detection

Sometimes KeyView detects certain file formats (such as CSV) as ASCII because of the content of the documents. In such cases, you can configure KeyView to use the file extension to determine the document format. Using the file extension can improve detection of formats such as CSV, but might not detect text files successfully if they have incorrect file extensions.

To use the file extension for ASCII files during detection

- In the `formats.ini` file, set the following flag in the `detection_flags` section:

```
[detection_flags]
use_extension_for_ascii=1
```

The default is 0 (do not use the file extension).

Allow Consecutive NULL Bytes in a Text File

By default, if a document contains consecutive NULL bytes, it is not detected as text. Depending on the type of files you are working with, changing the default might increase detection accuracy.

To allow consecutive NULL bytes of ASCII characters in text files

In the `formats.ini` file, set the following flag in the `detection_flags` section:

```
[detection_flags]
ascii_allow_null_bytes=1
```

The default value is 0 (do not allow consecutive NULL bytes).

Translate Format Information

Format information can include file attributes in the following categories:

- Major format
- File class
- Minor format
- Major version
- Minor version

Not all categories are required. Many formats only include major format and file class, or major format only.

The format information has the following structure:

```
MajorFormat.FileClass.MinorFormat.MajorVersion.MinorVersion
```

For example:

```
81.2.0.9.0
```

Each number in the format information represents a file attribute. The entry `81.2.0.9.0` represents a Lotus 1-2-3 Spreadsheet file version 9.0, where:

81 = Lotus 1-2-3 Spreadsheet (major format)

2 = Spreadsheet (file class)

0 = not defined (minor format)

9 = 9 (major version)

0 = 0 (minor version)

The example above applies to `formats_e.ini` file. When extracting format information by using the function `getAutoDetectInfo()` method, the same format information is represented as `294.2.0.9`.

NOTE: The format values returned by `getAutoDetectInfo()` differ from those in `formats_e.ini` because the former defines a unique ID for each major format, whereas the latter uses a major version, minor version, and minor format to distinguish between formats.

Distinguish Between Formats

The structure `getAutoDetectInfo()` method provides a unique ID for each major format. For example, a call to `getAutoDetectInfo()` returns `351.1.0` for a Microsoft Word 2003 XML format. The major format 351 is unique to this format.

Unlike `getAutoDetectInfo()`, the `formats_e.ini` file distinguishes between formats by using the major version number. For example, in `formats_e.ini`, a Microsoft Word 2003 XML format is defined as `285.1.0.100.0`. The major format 285 and file class 1 are the same values for generic XML. The major version 100 distinguishes the format as Microsoft Word 2003 XML.

The major version is used in `formats_e.ini` to specify the following formats:

- The Microsoft Office 2003 XML format has the same major format and file class as generic XML (285.1). It is distinguished from generic XML by using the following major versions:

- Word: 100
- Excel: 101
- Visio: 110
- The XHTML format has the same major format and file class as HTML (210.1). It is distinguished from HTML by using the major version 100.

Determine a Document Reader

The format detection module uses the `formats_e.ini` file to determine whether a format is supported and which reader should be used to parse a format. The entries in the `formats_e.ini` file lists each format's coded value, and an abbreviation for the format's reader. For example:

```
81.2.0.9.0=1123
```

The reader abbreviation is a truncated version of the reader's library name. Adding "sr" to the end of an abbreviation creates the name of the reader. The example entry above specifies that a Lotus 1-2-3 Spreadsheet file version 9.0 is parsed by the Lotus 1-2-3 reader, 1123sr.

[Files Required for Redistribution, on page 240](#) lists the document readers provided with KeyView.

Category Values in formats_e.ini

The [Detected Formats](#) section lists all of the file formats that can be detected by KeyView, with associated category values for use in the `formats_e.ini` file. The following tables provide the list of possible file classes and minor formats.

- [File Classes](#)
- [Minor Formats](#)

File Classes

| Attribute Number | Description | File class |
|------------------|----------------|-----------------|
| 0 | No file class | AutoDetNoFormat |
| 01 | Word processor | adWORDPROCESSOR |
| 02 | Spreadsheet | adSPREADSHEET |
| 03 | Database | adDATABASE |
| 04 | Raster image | adRASTERIMAGE |
| 05 | Vector graphic | adVECTORGRAPHIC |
| 06 | Presentation | adPRESENTATION |

File Classes, continued

| Attribute Number | Description | File class |
|------------------|------------------------|-----------------|
| 07 | Executable | adEXECUTABLE |
| 08 | Encapsulation | adENCAPSULATION |
| 09 | Sound | adSOUND |
| 10 | Desktop publishing | adDESKTOPPUBLSH |
| 11 | Outline/planning | adOUTLINE |
| 12 | Miscellaneous | adMISC |
| 13 | Mixed format | adMIXED |
| 14 | Font | adFONT |
| 15 | Time scheduling | adSCHEDULE |
| 16 | Communications | adCOMMUNICATION |
| 17 | Object module | adOBJECTMODULE |
| 18 | Library module | adLIBRARY |
| 19 | Fax | adFAXFORMAT |
| 20 | Movie | adMOVIE |
| 21 | Animation | adANIMATION |
| 22 | Source Code | adSOURCECODE |
| 23 | Computer-Aided Design | adCAD |
| 24 | BI and analysis tools | adANALYTICS |
| 25 | Scientific data | adSCIENTIFIC |
| 26 | Geographic Info System | adGIS |

Minor Formats

| Attribute Number | Minor Format |
|------------------|--------------------------|
| 00 | Minor format not defined |
| 01 | Standard |
| 02 | Book |

Minor Formats, continued

| Attribute Number | Minor Format |
|------------------|--------------------|
| 03 | Chart |
| 04 | Macro |
| 05 | Text |
| 06 | Binary |
| 07 | PC |
| 08 | Windows |
| 09 | DOS |
| 10 | Macintosh |
| 11 | RGB |
| 12 | TIFF |
| 13 | IFF |
| 14 | Experimental |
| 15 | Format Information |
| 16 | RLE |
| 17 | Symbol |
| 18 | Old |
| 19 | Footnote |
| 20 | Style |
| 21 | Palette |
| 22 | Configuration |
| 23 | Activity |
| 24 | Resource |
| 25 | Calculation |
| 26 | Glossary |
| 27 | Spelling |
| 28 | Thesaurus |

Minor Formats, continued

| Attribute Number | Minor Format |
|------------------|---------------|
| 29 | Hyphenation |
| 30 | Miscellaneous |
| 31 | UNIX |
| 32 | VAX |
| 33 | Driver |
| 34 | Archive |

Appendix G: Files Required for Redistribution

This section lists the Export files that can be redistributed in your applications under the licensing agreement. Unless noted, these files are in the directory *install\OS\bin*, where *install* is the path of the Export installation directory and *OS* is the operating system platform.

- [Core Files](#) 240
- [Support Files](#) 241
- [Document Readers and Writers](#) 243

NOTE: On Windows systems, the libraries are .dll files. On UNIX systems, the libraries are .so, .a, or .sl files.

Core Files

The following core files can be redistributed with your application.

| File | Description |
|---------------|---|
| formats_e.ini | Initialization file. For more information on this file, see Determine Format Support, on page 232 . |
| *htmlexport.* | Required by the Java API. |
| htmserv.dll | The in-process version of the HTML Export COM interface. |
| htmserv.exe | The out-of-process version of the HTML Export COM interface. |
| htmcnv.* | HTML converter for the document token stream. |
| KeyView.jar | Interface for Java support. NOTE: This file can be found at the path <i>install/javaapi/KeyView.jar</i> where <i>install</i> is the Export SDK installation directory. |
| kpifcnvt.* | Graphic conversion routines. |
| kpifutil.* | Graphic utility routines. |
| kvdecrypt.* | Decryption utility functions |
| kvxtract.* | File Extraction interface. |
| kvhtml.* | HTML Export C API. |
| kvexport.* | Export C API. Interface to the HTML and XML Export C APIs. |

| File | Description |
|------------------|---|
| kvexportdotnet.* | Interface for .NET support. |
| kvolefio.* | Embedded OLE object writer. |
| kvutil.* | Internal KeyView utility functions. |
| kvxpgsa.* | Interface between presentations or graphic readers and the Export API. |
| kvxssa.* | Interface between spreadsheet readers and the Export API. |
| kvxwpsa.* | Interface between word processing readers and the Export API. |
| kvzip.* | Zip writer |
| kwad.* | File auto-recognition module. |
| regsvr32.exe | A Microsoft Windows program used to register in-process COM objects. |
| txtcnv.* | Converter for document token stream. |
| *xmlexport.* | Required by the Java API. |
| *\vcredist* | (Windows platforms only) Microsoft Visual C++ Redistributable Packages. NOTE: This folder can be found in the Export SDK installation directory. |

Support Files

The following support files can be redistributed with your application.

| File | Description |
|------------------------|--|
| datafiles\ | (Folder) Required by kvlangdetect. |
| NSFtemplates\ | (Folder) Templates used by nsfsr to format Lotus mail notes. |
| 7z.* | Required by z7zsr and multiarcsr. |
| bentofio.* | Required by 1123sr.* and kpprzrdr.*. |
| cbmap.map | Character mappings for Adobe Portable Document Format (PDF). |
| CEBDLL.* | Required by cebsr. |
| chartbls.ux | Character mapping tables. |
| chmdll.* | Required by chmsr. |
| *codeidentifierplugin* | Required for source code identification. |

| File | Description |
|-----------------|--|
| DFECore.* | Required by cbsr. |
| Filter.* | Required by cbsr. |
| kp3dwrld.* | Required for 3D charts. |
| kpchtrdr.* | Required for all spreadsheets (chart support). |
| kpjavwrt.* | Java utility routines. |
| kpjpeg.* | JPEG file interchange format shared routines. |
| kppng.* | Portable Network Graphics (PNG) utilities. |
| kvlangdetect.* | Utility functions for language and character set detection. |
| kvxconfig.ini | Contains element extraction settings for source XML files. |
| kvgraph.* | Required for all spreadsheets (chart support). |
| kvpie.* | Required for all spreadsheets (chart support). |
| kvradar.* | Required for all spreadsheets (chart support). |
| kv.lic | Contains license information for KeyView products. This file is opened and validated when a KeyView API is used. |
| kv raster.class | Java program used to convert vector graphics on UNIX and Linux. |
| kvVector.class | Java applet used to convert vector graphics on UNIX and Linux. |
| kvvector.jar | Java applet used to convert vector graphics on UNIX and Linux. This must reside in the output directory. |
| langdetectext.* | Required by kvlangdetect.* |
| libey32.dll | (Windows platforms only) SSL utility functions used by KeyView mail format readers. |
| libpff.* | Required by pffsr. |
| libstlport.so.1 | (Solaris platforms only) Solaris Studio Redistributable. |
| oleaut32.* | Microsoft OLE Automation Controls. |
| olepro32.* | Microsoft OLE property support library. |
| servant.exe | Executable required for out-of-process conversions. |
| unzipjpg.* | Required for JPEG decompression. |
| wpmap.* | Extended character mapping for WordPerfect and Corel Presentation. |
| xmlsh.* | Contains a library of content handlers for each XML file type. Required by the Expat XML parser. |

Document Readers and Writers

The following readers and writers can be redistributed with your application.

| File | Description |
|-------------|--|
| ad1sr.* | AD1 Evidence file reader |
| afsr.* | ASCII reader |
| assr.* | Applix spreadsheet reader |
| awsr.* | Applix Words reader |
| bkfsr.* | Microsoft Backup File reader |
| bmpsr.* | Windows bitmap (BMP) reader |
| bzip2sr.* | Bzip2 reader |
| cabsr.* | Microsoft Cabinet format reader |
| cebsr.* | Founder Chinese E-paper Basic reader |
| chmsr.* | Microsoft Compiled HTML Help reader |
| csvsr.* | Comma-Separated Values reader |
| dbfsr.* | dBase Database reader |
| dbxsr.* | Microsoft Outlook Express DBX reader |
| dcasr.* | Document Content Architecture/Revisable Form Text (DCA/RFT) reader |
| difsr.* | Data Interchange Format reader |
| dmgsr.* | Mac Disk Copy Disk Image File reader |
| dw4sr.* | DisplayWrite 4 reader |
| dx1sr.* | Domino XML Language reader |
| em1sr.* | Microsoft Outlook Express (EML) reader. This is used to convert EML files when the MBX reader is not licensed. |
| emxsr.* | Legato EMailXtender archive (EMX) reader |
| encasesr.* | Expert Witness Compression Format (EnCase) v6 reader |
| encase2sr.* | Expert Witness Compression Format (EnCase) v7 reader |
| entsr.* | Microsoft Entourage Database Format reader |
| epubsr.* | Open Publication Structure eBook reader |

| File | Description |
|------------|--|
| foliosr.* | Folio Flat File reader |
| gifsr.* | Graphics Interchange Format (GIF) reader |
| gwfssr.* | GroupWise FileSurf reader |
| hl7sr.* | Health level7 reader (metadata only) |
| htmsr.* | HTML and XHTML reader |
| hwposr.* | Hangul 2002, 2005, 2007 reader |
| hwpsr.* | Hangul 97 reader |
| ichatsr.* | Apple iChat Log reader |
| icssr.* | Microsoft Outlook iCalendar reader |
| isosr.* | ISO-9660 CD Disc Image Format reader |
| iwss13sr.* | iWork 13 Numbers reader |
| iwsssr.* | Apple iWork Numbers reader |
| iwwp13sr.* | iWork 13 Pages reader |
| iwwpsr.* | Apple iWork Pages reader |
| jp2000sr.* | JPEG 2000 metadata reader |
| jpgsr.* | JPEG metadata reader |
| jtdsr.* | JustSystems Ichitaro reader |
| kpagrdr.* | Applix Presents reader |
| kpanirdr.* | Animated cursor reader |
| kpbmprdr.* | Windows Bitmap reader |
| kpbmpwrt.* | Windows Bitmap writer |
| kpcdrdr.* | Corel Draw |
| kpcgmrdr.* | Computer Graphics Metafile reader |
| kpcgmwrt.* | Computer Graphics Metafile writer |
| kpdcxrdr.* | DCX (fax) reader |
| kpDWGrdr.* | AutoCAD Drawing format reader |
| kpDXFrdr.* | AutoCAD Drawing Exchange format reader |

| File | Description |
|---------------|---|
| kpemfrdr.* | Enhanced Metafile reader |
| kpemfwrt.* | Enhanced Metafile writer |
| kpepsrdr.* | Encapsulated PostScript (EPS) reader |
| kpgflrdr.* | OmniGraffle Picture reader |
| kpgifrdr.* | Graphic Interchange Format (GIF) reader |
| kpgifwrt.* | Graphic Interchange Format (GIF) writer |
| kpicordr.* | Windows Icon reader |
| kpiwpgdrdr.* | Apple iWork Keynote reader |
| kpjbig2rdr.* | JBIG2 reader |
| kpjp2000rdr.* | JPEG 2000 reader |
| kpjpgdrdr.* | JPEG file interchange format reader |
| kpjpgwrt.* | JPEG file interchange format writer |
| kpnbmprdr.* | IBM Notes Bitmap reader (for embedded images in DXL files) |
| kpmacrdr.* | MacPaint reader |
| kpsmordr.* | Microsoft Office Drawing Objects (office 97, 2000, and XP) reader |
| kpodfrdr.* | Oasis Open Document Format presentation (ODP) reader |
| kpODArdr.* | AutoCAD reader (Windows only) |
| kpONErdr.* | Microsoft OneNote reader |
| kpoxdrdr.* | Open Office XML Diagram Graphics reader |
| kppdfrdr.* | Adobe Portable Document File (PDF) graphic-based reader |
| kppdf2rdr.* | High-fidelity Adobe Portable Document File (PDF) graphic-based reader |
| kpp40rdr.* | Microsoft PowerPoint PC 4.0 and PowerPoint Mac reader |
| kpp95rdr.* | Microsoft PowerPoint 95 reader |
| kpp97rdr.* | Microsoft PowerPoint 97 and higher reader |
| kppctrdr.* | Macintosh Quick Draw Picture (PICT) reader |
| kppcxrdr.* | PC Paintbrush (PCX) reader |
| kppdfrdr.* | Graphic-based Adobe Portable Document File (PDF) reader |

| File | Description |
|-------------|--|
| kppicrdr.* | Pictor PC Paint format (PIC) reader |
| kppngrdr.* | Portable Network Graphics (PNG) reader |
| kppngwrt.* | Portable Network Graphics (PNG) writer |
| kpppxrdr.* | Microsoft PowerPoint XML reader 2007 |
| kpprerdr.* | Lotus Freelance Graphics for Windows V2.0 reader |
| kpprzrdr.* | Lotus Freelance Graphics 96/97/98 reader |
| kprawrdr.* | ODA Internal Raster (RAW) Picture reader |
| kpsddrdr.* | StarOffice Draw / Impress reader |
| kpsdwrdr.* | Lotus Ami Pro Graphics reader |
| kpsgirdr.* | SGI RGB reader |
| kpshwrdr.* | Corel Presentations reader |
| kpsprdr.* | Shape Stream reader |
| kpsunrdr.* | Sun Raster reader |
| kptgardr.* | Truevision Targa reader |
| kptifdr.* | Tagged Image File Format (TIFF) reader |
| kpvsd2rdr.* | Microsoft Visio reader |
| kpvsdxrdr.* | Microsoft Visio 2013 reader |
| kpwg2rdr.* | WordPerfect Graphics 2 reader |
| kpwmfrdr.* | Windows Metafile reader |
| kpwmfwrt.* | Windows Metafile writer |
| kpwpgrdr.* | WordPerfect Graphics 1 reader |
| kpxfd1rdr.* | Extensible Forms Description Language reader |
| kvgzsr.* | GZIP reader |
| kvhqxsr.* | BinHex reader |
| kvzeesr.* | UNIX Compress reader |
| l123sr.* | Lotus 123 v96/97/98 reader |
| lasr.* | Lotus AMI Pro reader |

| File | Description |
|--------------|---|
| ltbenn30.dll | Lotus Word Pro support (supported on Windows x86 platform only) |
| ltscsn10.dll | Lotus Word Pro support (supported on Windows x86 platform only) |
| lwpapin.dll | Lotus Word Pro support (supported on Windows x86 platform only) |
| lwppann.dll | Lotus Word Pro support (supported on Windows x86 platform only) |
| lwpsr.dll | Lotus Word Pro reader (supported on Windows x86 platform only) |
| lzhsr.* | Microsoft Compression Folder reader |
| macbinsr.* | MacBinary reader |
| mbsr.* | Microsoft Word Macintosh reader |
| mbxsr.* | Mailbox (MBX) ¹ and Microsoft Outlook Express (EML) reader |
| mdbsr.* | Microsoft Access reader. |
| mhtsr.* | MIME HTML reader |
| mifsr.* | Adobe Maker Interchange Format reader |
| misr.* | Microsoft Word 2 reader |
| mp3sr.* | MP3 reader for metadata extraction |
| mppsр.* | Microsoft Project reader |
| msgsr.* | Microsoft Outlook (MSG) reader |
| mspubsr.* | Microsoft Publisher reader |
| msw6sr.* | Microsoft Works 6 and 2000 reader |
| mswsr.* | Microsoft Works V1 and 2 reader |
| multiarcsr | ARJ reader |
| mw6sr.* | Microsoft Word 95 reader |
| mw8sr.* | Microsoft Word 97, 2000, and XP reader |
| mwsr.* | Microsoft Word for DOS and Microsoft Write reader |
| mwssr.* | Microsoft Works Spreadsheet reader |
| mwxsr.* | Microsoft Word 2007 XML reader |

¹This reader is an advanced feature and is sold and licensed separately from KeyView Export SDK.

| File | Description |
|-----------|---|
| nsfsr.* | IBM Notes Database reader ¹ |
| oa2sr.* | Fujitsu Oasys reader |
| odfsssr.* | Oasis Open Document Format spreadsheets (ODS) reader |
| odfwpsr.* | Oasis Open Document Format word processing (ODT) reader |
| olesr.* | Embedded OLE object reader. |
| olmsr.* | Microsoft Outlook for Macintosh reader |
| onmsr.* | Legato EMailXtender Native Message reader |
| oo3sr.* | Omni Outliner reader |
| pdf2sr.* | Alternative Adobe Portable Document Format file (PDF) reader |
| pdfsr.* | Adobe Portable Document File (PDF) reader |
| pffsr.* | Microsoft Outlook Offline Storage File reader |
| pngsr.* | Portable Network Graphics (PNG) reader |
| pstsr.dll | Microsoft Outlook Personal Folders file MAPI-based reader (supported on Windows platform only) ² |
| pstnsr.* | Microsoft Outlook Personal Folders file native reader ³ |
| pstxsr.* | Microsoft Outlook Personal Folders file native reader ⁴ |
| qpssr.* | Quattro Pro spreadsheet reader |
| rarsr.* | RAR Archive reader |
| rtfsr.* | Microsoft Rich Text Format reader |
| skypesr.* | Skype log file reader |
| sosr.* | StarOffice/OpenOffice reader |
| starcsr.* | StarOffice Calc reader |
| starwsr.* | StarOffice Writer reader |

¹This reader is an advanced feature and is sold and licensed separately from KeyView Export SDK.

²This reader is an advanced feature and is sold and licensed separately from KeyView Export SDK.

³This reader is an advanced feature and is sold and licensed separately from KeyView Export SDK.

⁴This reader is an advanced feature and is sold and licensed separately from KeyView Export SDK.

| File | Description |
|------------|--|
| swfsr.* | Macromedia Flash reader |
| tarsr.* | Tape archive reader |
| tifsr.* | TIFF reader (metadata only) |
| tnefsr.* | Transfer Neutral Encapsulation Format reader |
| unihtmsr.* | Unicode HTML reader |
| unisr.* | Unicode reader |
| unzip.* | Zip file reader |
| uudsr.* | UUEncoding reader |
| vsdsr.* | Microsoft Visio reader |
| vcfsr.* | Microsoft Outlook vCard Contact reader |
| wkssr.* | Lotus 1-2-3 v2.0 through 5.0 reader |
| wosr.* | WordPerfect 5.x reader |
| wp6sr.* | WordPerfect 6.0 through 10.0 reader |
| wpmsr.* | WordPerfect for Macintosh reader |
| xlsbsr.* | Microsoft Office 2007 Excel Binary Format reader |
| xlssr.* | Microsoft Excel reader |
| xlsxsr.* | Microsoft Excel 2007 XML reader |
| xmlsr.* | Generic XML reader |
| xpssr.* | XML Paper Specification reader |
| xywsr.* | XYWrite reader |
| yimsr.* | Yahoo! Instant Messenger reader |
| z7zsr.* | 7-Zip reader |

Appendix H: Password Protected Files

This section lists supported password-protected container and non-container files and describes how to open them.

- [Supported Password Protected File Types](#)250
- [Export Password Protected Files](#) 251
- [Open Password Protected Container Files](#)251

Supported Password Protected File Types

The following table lists the password-protected file types that KeyView supports.

Key to support table

| Symbol | Description |
|--------|--|
| Y | Format is supported. |
| N | Format is not supported. |
| S | Support for viewing subfiles. |
| V | Support for viewing content. |
| P | Password required. |
| C | Password and certificate or User ID file required. |

Supported password-protected file types

| File Type | Version | Filter | Export | Extract | View | Credentials |
|--------------------------------|---------|--------|--------|---------|------|-------------|
| PST (Windows) | n/a | N | N | Y | S | P |
| PST (non-Windows) ¹ | n/a | N | N | Y | S | N |
| ZIP | n/a | N | N | Y | S | P |
| 7-Zip | n/a | N | N | Y | S | P |
| RAR | n/a | N | N | Y | S | P |
| SMIME in MSG, EML, MBX | n/a | N | N | Y | N | C |

¹The native PST readers, pstxsr and pstnsr, do not require credentials to open password-protected PST files that use compressible encryption.

Supported password-protected file types, continued

| File Type | Version | Filter | Export | Extract | View | Credentials |
|------------------|-------------------------|--------|--------|---------|------|-------------|
| Lotus Notes NSF | n/a | N | N | Y | N | C |
| Adobe PDF | n/a | Y | Y | Y | V | P |
| Microsoft Office | 97-2003 2007 2010 | Y | Y | Y | V | P |

Export Password Protected Files

This section describes how to export password-protected non-container files with the Java API.

To export password-protected files

1. Create an instance of the `ConfigOption` class, and set the `OptionType` argument to `CFG_SETPASSWORD`, the `OptionValue` argument to `TRUE`, and the `OptionData` argument to the source file password. The password is a null-terminated string of 255 or fewer characters (the final byte is null).
2. Call the `setConfigOption` method and pass the `ConfigOption` object.
3. Call a `convert` method. See the Javadoc in the directory `install\javaapi\javadoc`, where `install` is the path name of the Export installation directory.

Open Password Protected Container Files

This section describes how to extract password-protected container files by using the Java API. The following guidelines apply to specific file types.

- **Notes NSF files.** If you are running a Notes client with an active user connected to a Domino server, you must specify the user's password as a credential regardless of whether the NSF files you are opening are protected. This enables KeyView to access the Notes client and the IBM Notes API. If the Notes client is not running with an active user, KeyView does not require credentials to access the client.
- **PST files.** To open password-protected PST files that use high encryption (Microsoft Outlook 2003 only), you must use the MAPI-based PST reader (`pstsr`). The native PST readers (`pstxsr` and `pstnsr`) do not support files that use high encryption and return the error message `KVERR_PasswordProtected` if a PST file is encrypted with high encryption.

To open container files

- Set the credential information to an `ExtOpenDocConfig` object, and pass it to the `extOpenDocument` method.

For example:

```
dconfig = new ExtOpenDocConfig();  
odconfig.setPassword(m_password);  
extContextID = m_objExportFilter.extOpenDocument(inFile, odconfig);
```

Send documentation feedback

If you have comments about this document, you can [contact the documentation team](#) by email. If an email client is configured on this system, click the link above and an email window opens with the following information in the subject line:

Feedback on HTML Export SDK Java Programming Guide (Micro Focus KeyView 12.3)

Add your feedback to the email and click **Send**.

If no email client is available, copy the information above to a new message in a web mail client, and send your feedback to swpdl.idoldocsfeedback@microfocus.com.

We appreciate your feedback!