



PicseI ePAGE

PDF file format support



Copyright

Copyright © PicseL 2002

Neither the whole nor any part of the information contained in, or the product described in, this document may be used, sold, transferred, adapted or reproduced in whole or in part in any manner or form or in any media except with the prior written permission of the copyright holder.

The product described in this document is subject to continuous development and improvement. All particulars of the product and its use contained in this document are given by PicseL in good faith. However, the content of this document is provided "as is", without warranties of any kind with regard to its accuracy or reliability, and specifically excluding all implied warranties, for example of merchantability, fitness for purpose, title and non-infringement. In no event shall PicseL be liable for any special, indirect or consequential damages whatsoever resulting from loss of use, data or profits, arising out of or in connection with the use of this document. PicseL reserves the right to revise the document or withdraw it at any time without prior notice.

No licence is granted (nor any rights of whatsoever nature granted), express or implied, to any intellectual property by this document.

Throughout this document, the trade names and trademarks of some companies and products have been used. No such uses are intended to convey endorsement of or other affiliations with the document.

Document number PICSEL-ESF0004B

June 2002

<http://www.picseL.com/>

Introduction



On handheld computers, traditional applications typically rely on synchronization with a PC to convert content from the original document format into a simpler format used by the device. This creates at least three problems. Firstly, some visual integrity is inevitably lost in the conversion process from a rich format to a simpler one, resulting in documents that do not look exactly like the original. Secondly, the reliance on a PC for synchronization means that direct network access is difficult, and the device is severely hampered when untethered from the PC. Thirdly, traditional applications take a monolithic approach, converting only a single format so that separate applications are needed for each different document type.

Picse!'s ePAGE applications present a new concept in document viewing software. Unlike traditional approaches, the viewer can access native files, so synchronisation is not necessary and files can be grabbed straight off PC filing systems, flash cards and networks/internet without the need for pre-conversion. ePAGE deals directly with the original document in its native format, and aims to faithfully represent all of the features of the original. Moreover, the novel software architecture based on Picse!'s ePAGE technology enables multiple format types to be richly supported in a single application.

ePAGE achieves a breakthrough in its ability to interpret and render complex file formats. Many apparently straightforward formats have in the past encountered difficulties even when used in their original application software on a different platform, for example when going from a PC to a Macintosh. Sometimes these files assume characteristics of the computer on which they are used, such as the screen size or byte order, and such assumptions have to be unravelled when using them on a different device. To compound the challenge, some formats may contain proprietary features which are not revealed to the public, and which even have been forgotten within the originating company. Others, such as HTML, have written standards yet are still subject to differing interpretations as evidenced by the varying treatment of web pages in proprietary browsers.

ePAGE supports the most popular global file formats. The formats interpreted by ePAGE are richly expressive, containing not just text but sophisticated layout and rendering features, rich fonts, colour, images, tables, graphics and many other document features. Picsel is continuously developing its file format support to eventually cover every feature of the native file. With such a wealth of features across many document types, this is inevitably an ongoing process, with milestone releases of new functionality planned at periodic intervals. The approach involves researching the feature set most commonly found in real documents, building support early for the most frequently used elements, and ensuring these features are reproduced with total faithfulness to the original. The emphasis of ePAGE is on displaying rich content rather than on reproducing the document creation facilities of the original application.

This document describes the features supported in ePAGE. This level of support already covers the vast majority of characteristics that occur in day to day documents of this type, and the specific features are described with notes where appropriate. Those features planned for future implementation are also described, for completeness.

PDF

Adobe Portable Document Format is widely used for transferring printable documents to other people. By including text and graphics, along with embedded fonts and content security controls, it allows people to publish documents to each other with confidence that they will look right, but without giving away full control of the material.

PDF is particularly popular for marketing materials on web sites, but also for a wide range of other purposes. Its design also includes the later addition of annotations, digital signatures, and even more pages of content to existing documents.

Feature	Support	Notes
Accepts PDF 1.0 documents	Yes	
Accepts PDF 1.1 documents	Yes	
Accepts PDF 1.2 documents	Yes	
Accepts PDF 1.3 documents	Yes	
Accepts PDF 1.4 documents	Yes	
Document-configurable page size	Yes	
Multi-page documents	Yes	

Feature	Support	Notes
Page dimensions	Yes	Includes page rotation
Headers and footers	Yes	
Page break	Yes	
Page numbering	Yes	
Appended pages	Yes	
Text using Western character sets	Yes	
Text using Asian characters	Yes	See section on Character Maps
Bold text	Yes	
Italic text	Yes	
Underlined text	Yes	
Strikeout text	Yes	
Pre-installed TrueType fonts supported	Yes	OEM system integrators should include suitable outline fonts.
Font Matching	Yes	ePAGE automatically selects the most suitable font from those available in the system
Font Switcing (fallback)	Yes	Needed in order to display characters that the current font does not include.
Font Widths	Yes	Adjusts text using embedded table of character widths when exact matching font is not available
Embedded fonts: - Type 1 - TrueType	Yes	
Character Encodings	Yes	See Character Map section
Embedded character encodings	Yes	
Font size	Yes	
Superscript / Subscript	Yes	
Bullet lists	Yes	
Ordered / numbered lists	Yes	
Alignment of text: - left - centre - right	Yes	
Justified text	Yes	
Text line spacing	Yes	
Indentation	Yes	
Rotated text	Yes	
Columns of text	Yes	

Feature	Support	Notes
Tab stops <ul style="list-style-type: none"> - left - centre - right - decimal 	Yes	
Tables	Yes	
Table borders	Yes	
Table border styles	Partial	Supports <ul style="list-style-type: none"> - single solid line - double solid line
Table captions	Yes	
Background colours	Yes	
Charts and graphs	Yes	
Uncompressed bitmap images	Yes	
Compressed bitmap images <ul style="list-style-type: none"> - ASCIIHexDecode - ASCII85Decode - LZWDecode - Flate - RunLengthDecode - CCITTFaxDecode - DCTDecode 	Yes	
Transformed images	Partial	Supports image rotation by 90, 180, 270 degrees, flipped images
Vector line paths	Yes	
Line join styles <ul style="list-style-type: none"> - bevelled - mitred - round 	Yes	
Line end cap styles <ul style="list-style-type: none"> - butt - square - round - triangle (arrowhead) 	Yes	
Line colour	Yes	
Path Fill: solid colour	Yes	
Transformations on paths	Yes	
Monochrome and Colour content <ul style="list-style-type: none"> - DeviceGray - DeviceRGB - DeviceCMYK - CalGray* - CalRGB* - Indexed - DeviceN 	Most	* Calgray and CalRGB are treated as DeviceGray and DeviceRGB
Compressed data streams	Yes	
Clipping paths	Yes	

Feature	Support	Notes
Encryption (for Digital Signatures)	Yes	Supports 40 bit encryption (R=2, V=1 and R=3, V=2) using empty password
Compatibility sections	Yes	

Standard Fonts

PDF assumes that a set of 14 standard fonts are guaranteed to be available in the reader or viewer. These fonts are as follows.

Courier
Courier-Bold
Courier-Oblique
Courier-BoldOblique
Helvetica
Helvetica-Bold
Helvetica-Oblique
Helvetica-BoldOblique
Times-Roman
Times-Bold
Times-Italic
Times-BoldItalic
Symbol
ZapfDingbats

The inclusion of these standard fonts with the ePAGE PDF agent is left as an option to the OEM, because the fonts consume device memory and on constrained devices this may not be acceptable to the OEM.

Best performance will be obtained when the standard fonts are included in the device, thus ensuring exact reproduction of the text as intended by the document author who created the PDF original. The ePAGE PDF agent scans the device for all installed fonts and will make use of the standard fonts exactly as intended when they are available.

However the ePAGE PDF agent has been created primarily for mobile devices, and recognises that the standard fonts may not be available due to memory constraints. In this case, ePAGE will automatically find the font on the system that is the closest match to the specified standard font, and all of the text will be fully rendered in this matching font. Document content is never lost, and text is fully readable, but the effect of using a substitute font may cause slight differences in the text metrics. This can result in some text being slightly mispositioned compared to

the original, or text lines that are longer or shorter than those using the correct standard font. In practice, these discrepancies tend to be minor, but ultimately the choice whether to include the standard fonts must be made by the OEM. PicseL have designed the PDF agent to provide flexibility and a choice of either option.

Character Maps

ePAGE can work with all characters in the Unicode specification. PDF documents are often encoded in other encodings, or reference one of the character collections (Adobe-Japan1-2, Adobe-Korea-1, Adobe-CNC1-0, or Adobe-GB1-2) used to describe the major Asian languages. ePAGE handles these by detecting the character encoding used, and converting it to UTF-16 Unicode. This conversion is done using the scheme of CMap conversion tables created by Adobe for the PDF format. The CMap character mappings supported by ePAGE are shown in the following table. (In addition to the encoding support, the display of characters is also dependent on the availability of appropriate fonts installed on the device.)

Western character encoding CMaps

StandardEncoding	MacRomanEncoding
WinAnsiEncoding	MacExpertEncoding
PDFDocEncoding	

Japanese character encoding CMaps

83pv-RKSJ-H	
90ms-RKSJ-H	90ms-RKSJ-V
90msp-RKSJ-H	90msp-RKSJ-V
90pv-RKSJ-H	
Add-RKSJ-H	Add-RKSJ-V
EUC-H	EUC-V
Ext-RKSJ-H	Ext-RKSJ-V
UniJIS-UCS2-H	UniJIS-UCS2-V
UniJIS-UCS2-HW-H	UniJIS-UCS2-HW-V
Identity -H	Identity-V
H	V

Traditional Chinese encoding CMaps

Available from PicseL on request

<i>B5pc-H</i>	<i>B5pc-V</i>
<i>ETen-B5-H</i>	<i>ETen-B5-V</i>
<i>ETenms-B5-H</i>	<i>ETenms-B5-V</i>
<i>CNS-EUC-H</i>	<i>CNS-EUC-V</i>

<i>UniCNS-UCS2-H</i>	<i>UniCNS-UCS2-V</i>
----------------------	----------------------

Simplified Chinese encoding CMaps**Available from Picsel on request**

<i>GB-EUC-H</i>	<i>GB-EUC-V</i>
<i>GBpc-EUC-H</i>	<i>GBpc-EUC-V</i>
<i>GBK-EUC-H</i>	<i>GBK-EUC-V</i>
<i>UniGB-UCS2-H</i>	<i>UniGB-UCE2-V</i>

Korean CMaps**Available from Picsel on request**

<i>KSC-EUC-H</i>	<i>KSC-EUC-V</i>
<i>KSCms-UHC-H</i>	<i>KSCms-UHC-V</i>
<i>KSCms-UHC-HW-H</i>	<i>KSCms-UHC-HW-V</i>
<i>KSCpc-EUC-H</i>	
<i>UniKS-UCS2-H</i>	<i>UniKS-UCS2-V</i>

The Simplified Chinese, Traditional Chinese and Korean encodings are not included by default in the Picsel product range, but can be added on request.

Future Support

Picsel continuously strives to enhance the feature set of its software. In addition to the above features supported in Version 1.0, the following features are planned to be incorporated in future versions of the Picsel software. (The ordering in the list does not imply any measure of priority among the items.)

Dotted lines
Transparency
Path fill patterns
Arbitrary image transformations
Embedded Type 3 or Multiple Master fonts
Symbol and ZapfDingbats encoding and character sets
Sound, video playback
Document information (properties), Page information
Presentations: page duration, transitions
Pre-press features: OPI, JobTickets, Colour trapping and separations
Lab and ICC based colour
Annotations (text, action, links, etc)
Bookmarks (outline navigation)
Thumbnail navigation
Interactive Forms, JavaScript

