Picsel ePAGE

Word file format support

Copyright

Document number PICSEL-ESF-0002C

April 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.

Picsel'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 Picsel'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.

Word

Microsoft Word is a word processing application that has become the industry standard, as part of Microsoft Office. It employs a binary file format that has undergone several significant revisions coinciding with the release of new versions of the Office suite. As a fully fledged word processor, the format is mainly devoted to the support of very rich text layout and styling features. In addition it allows the inclusion of pictures, drawings and Clip Art in a variety of bitmap and metafile formats, such as Windows MetaFiles and JPEG photographs. Like other applications in the Office suite, it can contain links to OLE objects of other applications.

The level of support for Word features provided by ePAGE is shown in the following table.

Feature	Support	Notes
Supports Word 95 documents	Yes	The format employed in Word 95 is the MS Word 6.0 format

Feature	Support	Notes
Supports Word 97 documents	Yes	The format employed in Word 97 is the MS Word 8.0 format
Supports Word 2000 documents	Yes	
Fonts supported Makes use of outline Truetype fonts installed in the device	Yes	OEM system integrators should include suitable outline fonts.
Additional fonts	Yes	A wide range of commonly used fonts may be installed for use by ePAGE
Font matching	Yes	ePAGE automatically selects the most suitable font from those available in the system
Font switching (fallback)	Yes	Needed in order to display characters that the current font does not include.
Line breaking	Yes	Needed in order to break scripts written without spaces, eg Chinese, Korean and Japanese.
International Language support	Yes	Refer to the section on 'Character Sets' for more information.
Font Size. All font sizes are supported.	Yes	Although MS Word imposes upper and lower limits to a displayed font, in ePage, there is no theoretical limit.
Text colour	Yes	
Bold text	Yes	
Italic text	Yes	
Underline text	Yes	
Supports overline and strikethrough of text with variable line thickness. Option to apply lines to spaces between words.	Yes	
Bulleting	Yes	
Ordered/numbered lists	Yes	
Lists with nested counters	Yes	

Feature	Support	Notes

Text alignment:
   -    Left
   -    Centre
   -    Right
	Yes	

Indentation:
   -    First line
   -    Hanging
   -    Negative
	Yes	

Text line height spacing
   -    absolute size
   -    size relative to text size
   -    additional leading
	Yes	Includes single spacing, 1.5 and double line spacing, amongs others

Columns of text.	Yes	May be balanced or unbalanced, with column width and spacing specified as absolute or percentage distances.
Tabstops	Partial	Supports left tab stops and default tab stops
Tables.	Yes	
Tables with Indentation, Bullets, Numbering.	Yes	
Table borders	Partial	Tables may have borders around the whole table, groups of cells or individual cells.

Table border styles	Partial	Supports
		   - single solid line
		   - double solid line

Table Captions	Yes	Captions are placed next to the object they are associated with. Top and Bottom caption associations are supported.
Table – background cell colour	Yes	
Table text - horizontal alignment	Yes	Supports left, centre and right alignment of text within table cell.

Feature	Support	Notes
Page dimensions	Yes	Supports different page sizes of sections within same document
Page margins and borders	Yes	
Headers and Footers	Yes	
Page Break	Yes	
Page Number	Yes	
Supports "Next Page" section breaks in Word documents (where the new section starts on the next page)	Yes	
Cross References	Yes	Cross references are displayed, but are not navigable.
Index, Contents and Tables.	Yes	The TOC contents are displayed exactly as it is in the original document, even if the contents are moved.
Bookmark	Yes	. Bookmarks are translated, but are not viewable.
Document Views	Partial	ePAGE always opens up a document in 'print' view format.

Character Sets

ePAGE can work with all characters in the Unicode specification. Many documents are encoded in other encodings, and ePAGE handles these by automatically detecting the character encoding used, and converting it to UTF-16 Unicode. The character encodings 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.)

Encoding	Comment
ASCII	
UTF8	Unicode
UTF16	Unicode
Latin1	Western
Shift-JIS (Windows codepage 932)	Japanese
ISO-2022-JP	Japanese
EUC-JP	Japanese

GB 2312	Chinese
Big 5	Chinese
Windows codepage 874	Thai
Windows codepage 936	Simplified Chinese GBK
Windows codepage 949	Korean EUC-KR
Windows codepage 950	Chinese traditional Big5
Windows codepage 1250	Eastern European
Windows codepage 1251	Cyrillic
Windows codepage 1252	Latin
Windows codepage 1253	Greek
Windows codepage 1254	Turkish
Windows codepage 1255	Hebrew
Windows codepage 1256	Arabic
Windows codepage 1257	Baltic
Windows codepage 1258	Vietnamese

Future Support

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

Feature
Subscript and Superscript text
Shading behind text , highlighted text
Vertical and Horizontal justification
Additional table border styles
Sub-documents including Excel worksheets
Right and centre tab stops
Tab stops with added leader characters
Odd Page, Even Page & Continuous section breaks
Date and time fields within document body
Footnote and Endnote.
Dynamic adjustment of Index, Contents and Tables if the contents move.
Autotext.
Versions, Properties, Reviewing, Protect document
Watermarks.
Hyperlinks, Fill-in forms
Drawing: lines, curves and freeform drawings.
Drawing: textboxes, autoshapes, WordArt, 3D
Database tables.
Bitmap images such as JPEG, GIF, PNG, BMP
Metafile support including WMF, EMF, TIF (most Clip Art is metafile format)