Picsel ePAGE

HTML file format support

Copyright

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.

HTML

Hypertext Markup Language is used for the text and structure of web pages on the World Wide Web. It was originally designed to represent scientific papers in a form that allowed different organisations to print them in different styles, but has since been extended to cover formatting and page layout issues, and of course to include pictures.

HTML is easy to write. Even relatively non-technical people can write it directly into a text editor, including hypertext links to other pages, references to images, document formatting, colours, and other details. Unfortunately, decoding it is more difficult as there are many small errors which are interpreted in certain ways by popular software, and have become assumed to be part of the standard.

Feature	Support	Notes
HTML 3.2 and earlier	Yes	
HTML 4.0	Yes	Many extensions also supported.
Unclosed tags	Yes	
Tags in wrong context	Yes	
Invalid attributes	Yes	
Missing required tags	Yes	

Feature	Support	Notes
Deprecated / proprietary tags	Yes	
Paragraph alignment: - left - centre - right	Yes	
Vertical alignment: - top - middle - bottom	Yes	
Measurement units: - px - pt - en - em - in - cm - mm - %	Yes	
Named & measured colours	Yes	
Heading paragraph styles	Yes	
Specialised paragraph styles (address, citation, variable, etc)	Yes	
Inline style definitions using style attribute	Yes	
International character sets	Yes	See section below.
Bullet lists	Yes	
Numbered lists	Yes	
Definition lists	Yes	
Font face	Yes	
Font size	Yes	
Font colour	Yes	
Italic	Yes	
Bold	Yes	
Underline	Yes	
Strikethrough	Yes	
Emphasis	Yes	
Page background colour	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

Feature	Support	Notes
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.
Hypertext links	Yes	
Tables	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	Top and Bottom caption associations are supported.
Table – background cell colour	Yes	
Nested tables	Yes	
Fixed size table cells	Yes	
Dynamic size table cells	Yes	
International character encodings	Yes	
Static GIF images	Yes	
GIF interlacing	Yes	
GIF transparency	Yes	
JPEG images	Yes	
PNG images	Yes	
BMP images	Yes	

Feature	Support	Notes
Interactive Forms - check boxes - radio buttons - text entry fields - submit buttons - reset buttons - default values - submission - JavaScript behaviour	Yes	
Image maps (clickable location-sensitive graphics) - rectangular areas - circular areas - polygon areas	Yes	
CSS (Cascading Style Sheets)	Partial	

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 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.)

The emphasis in this version of Picsel ePAGE is on document viewing, while the features below introduce more powerful web browsing capabilities.

Superscript, subscript, marquee text
Text background colour / image
Dynamic HTML & Layers
Full CSS stylesheet support
XHTML
Frames & target windows
Animated bitmap images
Clickable Imagemaps