

クラウド時代のサーバ組込用 テキスト抽出エンジン

Text Porter V5 Server

ファイル識別機能

ファイルを作成したアプリケーション名称とバージョンを識別します。

ページ抽出機能

指定したファイルの指定ページを取り出します。

ストリーム出力機能

テキスト抽出、プロパティ抽出、ページ抽出機能の出力先ストリームを指定することができます。

テキスト抽出機能

指定したファイルからアプリケーション独自の制御コードを取り除き、テキスト文字列を取り出します。

プロパティ抽出機能

指定したファイルのプロパティ情報を取り出します。

暗号化ファイルの対応

従来のPDFの暗号化ファイルに加え、新たにMicrosoft Officeのパスワードで保護された暗号化ファイルからもテキスト抽出ができます。

(暗号化ファイルからテキストを抽出するには、APIよりパスワードを入力する必要があります)

テキスト符号化方式

- テキストの符号化方式は、自由に追加・拡張できる設計になっています。
- 標準では次の文字集合データ(XMLによる定義データ)を用意しています。

1 国際基準

- UTF-16 (BigEndian/LittleEndian)
- UTF-8
- ISO-10646-UCS-2
- ISO-10646-UCS-4

2 日本語

- Shift-JIS-2004
- ISO-2022-JP-2004
- EUC-JIS-2004
- Shift-JIS
- Windows31J
- EUC-JP
- ISO-2022-JP

3 英語

- ISO-8859-1

4 中国語・韓国語

- GB18030
- GBK
- BIG5
- KS_C_5601_1987

利用例

- 全文検索エンジンと共にグループウェアに組み込み
- 企業内文書管理システムに組み込み
- 電子メール・サーバ製品に組み込み
- 文書管理ソフトウェア製品に組み込み
- インターネット情報検索アプリケーションに組み込み
- イントラネット・テキスト・マイニング・サーバに組み込み
- セキュリティ管理製品に組み込み

テキスト抽出対象ファイル

アプリケーション		Japanese	English	GB18030	GBK	BIG5	Korean	OLE IN	OLE OUT
Microsoft	Word Ver6/Ver7(95)/97/98/2000/XP/2003/2001 for Macintosh/XP/2003	○	○	○	○	○	○	○	○
	Word 2007/2010/2013/2016/2019	○	○	○	○	○	○	○	○
	Excel Ver4/Ver5/Ver7(95)/97/98 for Macintosh/2000/2001 for Mac/XP/2003	○	○	○	○	○	○	○	○
	Excel 2007/2010/2013/2016/2019	○	○	○	○	○	○	○	○
	PowerPoint 95/97/2000/XP/2003/2001 for Mac/XP/2003	○	○	○	○	○	○	○	○
	PowerPoint 2007/2010/2013/2016/2019	○	○	○	○	○	○	○	○
	RTF	○	○	○	○	○	○	○	○
	Works 2000 WP(文書)	○	—	—	—	—	—	○	—
	Works 2000 SS(表)	○	—	—	—	—	—	—	○
	Works 2000 DB(データベース)	○	—	—	—	—	—	○	—
	Visio 2000/2002/2003/2007/2010	○	—	—	—	—	—	○	○
	Visio 2013	○	○	○	○	○	○	○	○
	Outlook 2000/2003/2007/2010(*.msg形式)	○	○	○	○	○	○	○	—
	Outlook Express 6(*.eml形式)	○	○	○	○	○	○	○	—
Mozilla	Thunderbird 60.8.0(*.eml形式)	○	○	—	—	—	—	—	—
Visio	Visio V4/V5	○	—	—	—	—	—	○	○
ジャストシステム	一太郎 Ver5/Ver6/Ver6.3/Dash2	○	—	—	—	—	—	—	—
	一太郎 7	○	—	—	—	—	—	○	—
	一太郎 Lite、8から13、2004から2019	○	—	—	—	—	—	○	○
Adobe	Adobe PageMaker 6.0/6.5/7.0	○	○	○	○	○	○	○	—
PDF	PDF 1.2から1.7	○	○	○	○	○	○	—	○
富士通	OASYS V3/V4/V5/V6/V7/V8/2002/V10	○	—	—	—	—	—	—	—
	OASYS Online形式	○	—	—	—	—	—	—	—
Lotus	Lotus1-2-3 R5/97/98/OASYS 1-2-3 V6/V7/V8/2000	○	○	○	○	—	○	—	—
	Lotus1-2-3 MillenniumEdition9.5	—	○	—	—	—	—	—	—
Claris	Mac Write II	○	—	—	—	—	—	—	—
	ClarisWorks 4.0	○	—	—	—	—	—	—	—
Apple	AppleWorks 6	○	○	—	—	—	—	—	—
Corel	WordPerfect Office 2000(WordPerfect 8/9 のみ)	○	—	○	○	—	—	—	—
	Corel Presentations 9(Slide show 7/8/9)	○	○	—	—	—	—	—	—
Quark	QuarkXPress 3.3/4	○	—	—	—	—	—	—	—
AutoDesk	AutoCAD R13(LT95)/R14(LT97) DXF Ascii形式	○	○	○	○	○	○	○	—
	AutoCAD R13(LT95)/R14(LT97) DXF Binary形式	○	○	○	○	○	○	○	—
	AutoCAD 2000から2012 DXF	○	○	○	○	○	○	○	—
	AutoCAD R13(LT95)/R14(LT97) DWG	○	○	○	○	○	○	○	○
	AutoCAD 2000から2012 DWG	○	○	○	○	○	○	○	○
	IGES	○	○	—	—	—	—	—	—
富士フイルムビジネスソリューション	DocuWorks v4/v5/v6/v7/v8/v9	○	○	—	—	—	—	○	—
OpenOffice.org	OpenOffice.org 1.0	—	—	—	—	—	—	—	—
	OpenOffice.org 3.1/3.2/3.3, Libre Office 3.3/3.4	○	○	○	○	○	○	—	○
金山ソフト	WPS Office 97/2000/2002/2003 (金山文字のみ)	—	—	○	○	○	—	○	○
その他	HTML	○	○	○	○	○	○	—	—
	XML	○	○	○	○	○	○	—	—

※ PDF、Microsoft Office以外の、暗号化されたファイルからのテキスト及び、プロパティ抽出するには、予めオリジナルのアプリケーションで設定を解除する必要があります。
 ※ MicrosoftのIRM機能で、ドキュメントへのアクセス制限(閲覧、変更etc.)を設定したMicrosoft Officeファイルからのテキスト抽出はできません。
 ※ 表やデータベースファイルは、CSVファイルとして出力します。
 ※ LZH、ZIP、RARの3種類の圧縮ファイル(アーカイブファイル)については、表にあるテキスト抽出対象ファイルが含まれていた場合に出力可能です。

対応インターフェース

- COMインターフェース ● .NET インターフェース
 - Javaインターフェース ● .NET 5.0 (旧名 .NET Core) インターフェース
- ※ プラットフォームによって、サポートしていないものもあります。

動作環境

- Windows 10(32bit/64bit)
- Windows Server 2016(64bit)/Windows Server 2019(64bit)
- Linux(64bit)

ライセンス形態

サーバライセンス

サーバ上でプログラムを稼働し、ネットワークに接続したクライアントに機能を提供するためのライセンスです。クライアントは同一企業内であることが原則です。

開発ライセンス

SIやコンサルタントなどの組織、システム開発担当者が、本製品を組み込んだシステムの開発・試験などの用途にのみ使用できるライセンスです。実運用は許諾していません。

OEMライセンス

本製品をお客様の製品に組み込み、契約規模内の本数で複製、再頒布の権利を提供します。価格は契約規模に応じて算出します。営業窓口にご相談ください。

ASPライセンス

ASP/SaaSに利用するためのライセンスです。価格はサービス形態により個別算出します。営業窓口にご相談ください。

Text Porter について、詳しい情報はWebでご確認ください。 <https://www.antenna.co.jp/axx/>

■ 製品に関するお問い合わせ (土日祝日、年末年始を除く、平日9:30~18:00)

アンテナハウス株式会社 システム営業グループ

TEL : 03-5829-9021 e-mail : sis@antenna.co.jp

インターネットで詳しい情報を提供しています。

<https://www.antenna.co.jp/>



〒103-0004 東京都中央区東日本橋2-1-6 東日本橋藤和ビル5階
©2022 AntennaHouse, Inc.